

**Investigation of *de novo* methylation activity  
in mutants of the *Eco*KI methyltransferase**

Markus Winter

Thesis presented for the degree of Doctor of Philosophy

Institute of Cell and Molecular Biology

University of Edinburgh

1997





Come, my friends,  
'Tis not too late to seek a newer world.  
... and tho'  
We are not now that strength which in old days  
Moved heaven and earth; that which we are, we are;  
One equal temper of heroic hearts,  
made weak by time and fate, but strong in will  
To strive, to seek, to find, and not to yield.

*Alfred Lord Tennyson*

To my teachers,

my mother who taught me how to care,

my father who taught me to tolerate,

Patricia A. McKillip for the riddle of my life,

and Regina, who taught me trust and betrayal



## **Acknowledgements**

First and foremost I would like to thank my supervisor, Noreen Murray, for the opportunity to do my PhD in her laboratory, her PostDoc Dave Dryden for his tutelage, and my second supervisor Richard Hayward for many friendly words.

I would also like to thank all the people who helped me with techniques or services, especially Laurie Cooper, Gareth King, Lynn Powell, Peter Thorpe, and Annette Titheradge.

I'm grateful to all those people I had the pleasure to meet - Damion, Catherine, Vicky, Diane, Mary, Jules, John, Glenn, Lucy, Kristina, Petra, Ruth, Laura, Pam, David, John D. and all the others, too numerous to mention.

And finally a very special thank you to David Finnegan, whose moral support was invaluable.



# Abstract

A group of mutants of the *EcoKI* R/M-system displaying de novo methylation activity have been isolated (Kelleher *et al.*, 1991). The mutated genes were transferred into an overexpressing plasmid vector. Two of the over-expressed proteins were purified to near homogeneity from clones transformed with the plasmids.

Cofactor binding activities of wild-type and the two mutant enzymes were compared by 1,8-anilino-napthalene sulphonic acid fluorescence displacement experiments. A DNA methylation assay based upon the transfer of a tritiated methyl group from the cofactor AdoMet to the substrate DNA was established and used to examine the dependency of the reaction on cofactor, substrate, and enzyme concentration. In addition the stability of the trimeric enzyme at different protein concentrations was followed by HPLC gel filtration.

Sequence alignments, secondary structure predictions, and tertiary structural modelling were used to show the similarity of the Type I system *EcoKI* with methyltransferases from other classes (especially Type II methyltransferases), thereby establishing a structural and suggesting an evolutionary link between the different methyltransferase classes. The information obtained by these comparisons enabled the subsequent modelling of a more refined model of the *EcoKI* structure (Dryden, D. T. F., Sturrock, S. S., and Winter, M., 1995. *Structural modelling of a type I DNA methyltransferase*. Nature Struct. Biol. 2, 632-635).

A model is proposed to explain the different activities observed in wild-type and mutant enzymes based on the biochemical and structural data obtained during these investigations.



## List of Abbreviations

A	adenine or adenosine
aa	amino acid
AdoHcy	S-adenosyl-L-homocysteine
AdoMet	S-adenosyl-L-methionine
AMPS	ammonium persulphate solution
ANS	(1, 8-anilino-naphthalene sulphonic acid)
APS	ammonium persulphate solution
ATP	adenosine-5'-triphosphate
bp	base pair(s)
BSA	bovine serum albumin
C	cytosine or cytidine
C-	carboxyl-
Ci	Curie
cm	centimetre
CM	carboxymethyl
CPK	colour scheme based upon the colours of plastic spacefilling models developed by Corey, Pauling and later improved by Kultun.
cpm	counts per minute
CTP	cytidine-5'-triphosphate
dam	DNA adenine methylation
dcm	DNA cytosine methylation
ddNTP	dideoxynucleotide-5'-triphosphate
DEAE	diethylaminoethyl
dH <sub>2</sub> O	distilled water
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide-5'-triphosphate
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
e.g.	Latin <i>exempli gratia</i> , for example
EMBL	European Molecular Biology Laboratory
ENase	DNA endonuclease
e.o.p.	efficiency of plating
frac	fraction



g	gram
g	relative centrifugal force
G	guanine or guanosine
GTP	guanosine-5'-triphosphate
$^3\text{H}$	tritium
H-bonds	hydrogen bonds
HCl	hydrochloric acid
HPLC	high performance (high pressure) liquid chromatography
hr	hour
hsd	host specificity determinant (host specific defence)
i.e.	latin <i>id est</i> , that is
IPTG	isopropyl- $\beta$ -D-thiogalactopyranoside
kb	kilobase(s); kilobase pairs
kD	kilodalton
l	litre
m	milli ( $10^{-3}$ )
M	molar
MCS	multiple cloning site
MES	2-[N-morpholino]ethanesulphonic acid
min	minute
ml	milliliter
mm	millimetre
m.o.i.	multiplicity of infection
mol	mole
MTase	DNA methyltransferase
MW	molecular weight
n	nano ( $10^{-9}$ )
N-	amino-
n.a.	not applicable
n.d.	not determined
nm	nanometre
nt	nucleotide
NTP	nucleotide-5'-triphosphate
O.D.	optical density
o/n	over night
-OH	hydroxyl group



<i>orf</i>	open reading frame
PAGE	polyacrylamide gel electrophoresis
PEG	polyethylene glycol
pfu	plaque forming units
pH	$-\log_{10}[\text{H}^+]$
phage	bacteriophage
PMSF	phenylmethylsulphonyl fluoride
PNK	polynucleotide kinase
R/M	restriction and modification
RBS	ribosome binding site
r.m.s.d.	reduced mean standard deviation
SAM	synonym for AdoMet
SDS	sodium dodecyl sulphate
sec	second
SSC	standard saline citrate
T	thymine or thymidine
TBE	Tris-borate-EDTA (buffer)
TE	Tris-EDTA (buffer)
TEMED	N, N, N', N'-teramethyl ethylene diamine
TRD	target recognising domain
Tris	2-amino-2-(hydroxymethyl)-1,2-propandiol
u	unit
UV	ultraviolet
UWGCG	University of Wisconsin Genetics Computer Group
V	Volt
v/v	volume : volume ratio
W	watt
w/v	weight : volume ratio
$\Delta$	deletion
$\epsilon$	extinction coefficient
$\mu$	micro ( $10^{-6}$ )
$\mu\text{F}$	microfaraday
$\Omega$	Ohm
$^{\circ}\text{C}$	degree centigrade



## Conventions

1. Bacterial genes are in italics and their products start with a capital letter, *e.g.* the *hsdR* gene encodes HsdR. Bacteriophage genes occasionally follow different conventions, *e.g.* the *cI* gene.
2. The nomenclature for restriction and modification (R/M) systems is according to Smith and Nathans (1973), where R. and M. followed by the enzymes name indicate the restriction endonuclease and methyltransferase respectively, *e.g.* R.*EcoRI* and M.*EcoRI*.
3. Prophages are indicated in brackets after the host, *e.g.* BL21(DE3).
4. The host on which a bacteriophage was last propagated is indicated after the phage symbol, *e.g.*  $\lambda$ .C600,  $\lambda$ .K, and  $\lambda$ .0 indicate, respectively, that bacteriophage  $\lambda$  was grown on C600, on an *EcoKI*-modifying strain, and on a strain with no R/M-system.
5. Restriction and modification phenotypes are indicated with superscript negative or positive symbols, *e.g.*  $r^-m^+$  denotes a restriction-deficient modification-proficient strain. Subscript letters are used to indicate the R/M-system associated with the phenotype, *e.g.*  $r_K^-m_K^+$  means restriction-deficient modification-proficient for *EcoKI*.
6. *EcoR124II* was formerly known as *EcoR124/3*.



## Amino Acids

Name	Three Letter Code	Single Letter Code
alanine	Ala	A
arginine	Arg	R
asparagine	Asn	N
aspartate	Asp	D
cysteine	Cys	C
glutamate	Glu	E
glutamine	Gln	Q
glycine	Gly	G
histidine	His	H
isoleucine	Ile	I
leucine	Leu	L
lysine	Lys	K
methionine	Met	M
phenylalanine	Phe	F
proline	Pro	P
serine	Ser	S
threonine	Thr	T
tryptophan	Trp	W
tyrosine	Tyr	Y
valine	Val	V



**The genetic code**

1 <sup>st</sup>	2 <sup>nd</sup>	U	C	A	G	3 <sup>rd</sup>
U		Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr ochre amber	Cys Cys opal Trp	U C A G
C		Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C A G
A		Ile Ile Ile Met	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G
G		Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	U C A G



**CHAPTER 1**

INTRODUCTION .....1

1.1. The phenomenon of restriction .....2

1.2. Restriction / Modification Systems.....4

1.3. Classification of Restriction / Modification Systems.....7

1.3.1. according to complexity and cofactor requirements.....7

1.3.1.1. Type II.....7

1.3.1.2. Type IIs.....7

1.3.1.3. Type III.....8

1.3.1.4. Type I.....8

1.3.2. according to the type of methylation.....12

1.4. Mechanism of Restriction and Modification.....16

1.4.1. Type II enzymes .....16

1.4.1.1. Endonucleases .....17

1.4.1.2. Methyltransferases .....21

        C5-Methyltransferases .....21

        A6-Methyltransferases .....31

        C4-Methyltransferases .....35

1.4.2. Type IIs enzymes .....36

1.4.3. Type I enzymes .....39

1.4.3.1. The subunits .....39

        HsdS .....39

        HsdM .....44

        HsdR.....45

1.4.3.2. The Methyltransferase .....46

1.4.3.3. The EcoKI system.....48

1.5. Biology and Evolution of R/M-Systems.....51



**CHAPTER 2**

MATERIALS & METHODS .....54

2.1. Bacterial Strains .....55

2.2. Phage Strains .....56

2.3. Plasmids.....57

2.4. Oligonucleotides.....58

2.5. Enzymes and Chemicals.....59

2.6. Media and materials .....60

2.7. Standard Solutions .....61

2.7.1. Standard solutions used for microbial techniques.....61

2.7.2. Standard solutions used for DNA preparations .....61

2.7.3. Standard solutions used for Hybridization.....61

2.7.3.1. for Colony Lifts .....62

2.7.3.2. for Hybridization .....62

2.7.3.3. for Digoxigenin detection.....62

2.7.4. Standard solutions used for DNA agarose gel electrophoreses.....63

2.7.5. Standard solutions used during protein purification .....63

2.7.6. Standard solutions used for PAGE .....63

2.7.6.1. Standard solutions used for non-denaturing PAGE.....63

2.7.6.2. Standard solutions used for SDS-PAGE .....64

2.7.7. Standard solutions used for gel retardation .....64

2.7.8. Standard solutions used for ANS assay.....64

2.7.9. Standard solutions used for HPLC filtration .....64

2.7.10. Standard solutions used for Methyltransferase assay .....65



2.8. Microbial Techniques .....	66
2.8.1. Growth of E. coli bacterial cultures.....	66
2.8.2. Storage of bacterial strains.....	66
2.8.3. Preparation of plating cells and bacterial lawns .....	66
2.8.4. Preparation of l phage plate lysates .....	67
2.8.5. Phage Titration .....	67
2.8.5.1. Spot Tests .....	67
2.8.5.2. Plate titrations.....	67
2.8.6. Preparation and Transformation of Competent Cells .....	68
2.8.6.1. CaCl <sub>2</sub> method.....	68
2.8.6.2. Electroporation.....	69
2.8.7. Detection of positive Transformants.....	69
2.8.7.1. Colony Lifts.....	69
2.8.7.2. Hybridization .....	70
2.8.7.3. Digoxigenin detection.....	71
2.9. DNA Techniques .....	72
2.9.1. Basic Techniques .....	72
2.9.1.1. Phenol / Chloroform extraction.....	72
2.9.1.2. Precipitation of DNA with Ethanol and Sodium Acetate.....	73
2.9.2. Preparation of plasmid DNA .....	73
2.9.2.1. Alkaline Mini-Prep .....	73
2.9.2.2. Boiling Method .....	74
2.9.2.3. QIAprep™ Plasmid Preparation .....	74
2.9.2.4. Large scale preparation of plasmid DNA .....	75
2.9.3. DNA modifications .....	77
2.9.3.1. Restriction Endonuclease Digestion of DNA.....	77
2.9.3.2. Phosphorylation of DNA fragments .....	77
2.9.3.3. Creating "blunt-ended" DNA fragments.....	78
2.9.3.4. Ligation of DNA.....	78



2.9.3.5. End Labelling of Oligonucleotides .....	78
2.9.3.6. Preparation of digoxigenin DNA probes .....	78
2.9.3.7. Annealing of complementary oligonucleotides.....	79
2.9.3.8. Site-directed mutagenesis by PCR .....	79
2.9.4. Analysis of DNA fragments on and purification from agarose gels .....	80
2.9.4.1. Agarose Gel Electrophoresis.....	80
2.9.4.2. DNA Fragment Preparation by Electroelution .....	81
2.9.4.3. DNA Fragment Preparation from Low Melting Point Agarose Gels.....	81
2.9.4.4. DNA Fragment Preparation with GeneClean II™.....	81
2.10. Protein Techniques .....	83
2.10.1. Expression of hsdMS.....	83
2.10.2. Protein Purification and Storage.....	83
2.10.3. Polyacrylamide Gel Electrophoresis (PAGE).....	85
2.10.3.1. Non-denaturing PAGE.....	85
2.10.3.2. SDS-PAGE.....	85
2.10.3.3. Staining of proteins with Coomassie .....	85
2.10.4. Desalting of protein .....	86
2.10.5. Determination of protein concentration in solution .....	86
2.11. Assays.....	87
2.11.1. Gel retardation assays.....	87
2.11.1.1. DNA binding assay.....	87
2.11.1.2. Dissociation assay .....	88
2.11.2. AdoMet Binding by Fluorescence of the extrinsic Fluorophore ANS .....	88
2.11.3. HPLC filtration .....	89
2.11.4. Methyltransferase assays .....	90
2.12. Software used .....	91



## CHAPTER 3

CLONING AND EXPRESSION.....	92
3.1. Introduction .....	93
3.2. Subcloning of m* mutants .....	94
3.2.1. Subcloning of hsdMS from $\lambda$ phage into the expression vectors pJF118HE or pJF118EH.....	94
3.2.2. Construction of N- and C-terminal deletion derivatives.....	96
3.2.3. Identification of positive clones .....	99
3.3. Expression from the constructs.....	99
3.4. Purification.....	102

## CHAPTER 4

BIOCHEMICAL CHARACTERISATION.....	106
4.1. AdoMet Binding by Fluorescence of an extrinsic Fluorophore.....	107
4.2. <i>In vitro</i> Methylase Activity Assay (by fluorographic Detection of [ <i>methyl</i> - $^3\text{H}$ ] Transfer).....	123
4.3. Assessment of the stability of the methyltransferase complex at low enzyme concentrations by HPLC .....	148

## CHAPTER 5

SEQUENCE ALIGNMENTS.....	151
5.1. Preface.....	152



5.2. Sequence Alignments .....	153
5.2.1. Limitations of alignment algorithms .....	153
5.2.2. Preparation of sequences.....	154
5.2.3. Alignments by Computer .....	155
5.2.4. Alignments by Hand.....	155
5.2.4.1. g-class .....	156
5.2.4.2. b-class .....	160
5.2.4.3. a-class.....	164
5.2.4.4. <i>EcoRI</i> and homologues.....	168
5.2.4.5. RNA methyltransferases.....	170
5.2.5. Comparison with structure based sequence alignments .....	173
5.2.6. Conclusion.....	174
5.3. Secondary Structure Predictions.....	174
5.3.1. Chou-Fasman .....	174
5.3.2. PredictProtein .....	175
5.4. Model Building .....	177
5.5. How do other data fit into the model?.....	181
5.6. Proposing a new structure for M. <i>EcoKI</i> .....	184

## CHAPTER 6

CONCLUSIONS AND GENERAL DISCUSSION .....	187
--	-----

References.....	199
-----------------	-----



# **CHAPTER 1:**

## **INTRODUCTION**



### 1.1. The phenomenon of restriction

Viruses exhibit extensive adaptability to growth in various hosts or tissues, and it was widely held in the past that virus adaptability reflected a peculiar plasticity of virus heredity, which allowed it to be directly influenced by its host cells (as quoted in: Luria, 1953). However in some instances selection of spontaneous mutants had been demonstrated to be the mechanism causing variation (Luria, 1945), and it became generally accepted that most variations in virus properties are caused by viral mutations, and that virus plasticity results from the variety of genotypes present in large viral populations.

As a consequence of this, one of virology's most generally valid rules until 1952 was that the properties of virus particles are unaffected by the hosts in which they grow. Host adaptation and tissue adaptation, the apparent exceptions, were explained by selective reproduction of virus mutants in new hosts or in new tissues (Findlay, 1939; Burnet, 1946).

In analysing the relation between certain coliphages and certain strains of their bacterial hosts, a transient change - with an efficiency of change too high to be explainable in terms of mutation and selection - was found (Luria and Human, 1952; Bertani and Weigle, 1953). It emerged that one cycle of growth in a suitable host affects the phenotype of the new virus: the phenotypic change suppresses (or enhances) the ability of the virus to reproduce in certain hosts therefore "restricting" (or enlarging) its host range, and the change is transient in the sense that one cycle of growth in a suitable host returns the virus to its original form. Therefore its characteristics were distinctly different from mutations in that it is strictly phenotypic, non-hereditary, and that it is determined by the host cell in which a virus has been produced. Other instances of host-controlled variation in bacteriophages were reported for *Staphylococcus* phages (Ralston and Krueger, 1952; Rountree, 1956), *Salmonella* phages (Anderson and Felix, 1952), and *Streptococcus* phage (Collins, 1956), and the phenomenon was renamed "host controlled restriction and modification" to distinguish it from genetic mutation (Arber, 1965b).

In a series of elegant experiments Arber *et al.* (1962) and Dussoix *et al.* (1962) showed that host controlled modification is carried and serially transferred (in either conserved or semi-conserved form) on the bacteriophage DNA, and that restriction is



achieved by the degradation of unmodified phage DNA upon infection of the new host strains (Arber, 1962; Arber and Dussoix, 1962; Dussoix and Arber, 1962; Arber *et al.*, 1963). It was also noted that the host controlled modification is governed both by the bacterial genome and by episomes which are present in the cell since in some instances the controlling loci could be mapped on the bacterial chromosome (Zinder, 1960), while in others they were known to be contained in the genome of an unrelated prophage (Lederberg, 1957) or resistance factor (Watanabe *et al.*, 1964; Arber and Morse, 1965). Based on this and preliminary evidence that bacterial DNA is also subject to this same control mechanism, Arber concluded that the modification would not only be imparted on phage DNA, but would be carried on the DNA of the host itself as well (Arber, 1962), as was subsequently shown by its effect on both transduction and bacterial conjugation (Boyer, 1964; Pittard, 1964; Arber and Morse, 1965).

The separation of restriction and modification into two independent functions both physiologically (Christensen, 1961; Drexler and Christensen, 1961; Arber *et al.*, 1963; Kellenberger *et al.*, 1966) and genetically (through the identification of *r-m*<sup>+</sup> mutants in phage P1 (Glover *et al.*, 1963) and later in the *Escherichia coli* strains K12 and B (Wood, 1965)) led Arber to propose that host specificity is mediated by the combination of a "restriction enzyme" and a "modification enzyme", both of which recognise the same specific base sequence (Arber, 1965b).

The nature of the modification was first suggested by Gold and Hurwitz who demonstrated in vitro that there are DNA methyltransferases in bacteria which methylate cytosine to 5-methylcytosine and adenine to 6-methyladenine and that DNA of different bacterial strains have unique methylation patterns. They therefore proposed that different methylation of lambda DNA by different host cells might confer different host specificities to the phage (Gold and Hurwitz, 1963; Gold *et al.*, 1963a; Gold *et al.*, 1963b). The issue was clouded somewhat by the fact that Ledinko (1964) found no increase in the level of 5-methylcytosine in DNA extracted from phage lambda carrying the host specificities derived from the strains *E.coli* C, B, K, or K lysogenic for phage P1, and that the presence of the majority of methylated bases (both 5-methylcytosine and 6-methylaminopurine) was found to be irrelevant to the specificity of host modification of DNA (Gough and Lederberg, 1966; Mamelak and Boyer, 1970) - due to the action of the then unknown *dcm* (DNA cytosine methylation) methylase and *dam* (DNA adenine methylation) methylase (which accounts for more than 99% of the 6-methyladenine present in *Escherichia coli*



(Marinus and Morris, 1973)). On the other hand the idea had received support by experiments with met<sup>-</sup> strains of *Escherichia coli* which showed that deprivation of methionine (but not of proline or arginine) during the latent growth period of the host cells results in the production of phage with a lower efficiency of plating on restricting host strains, and suggested that methionine is specifically required for the production of host specificity of DNA (Arber, 1965a; Grasso and Paigen, 1968; Grasso and Paigen, 1968). Following up on this, Arber found a correlation between occurrence of modification and the level of 6-methylaminopurine in phage DNA (Arber and Smith, 1966). Klein finally showed that T1-DNA may in some cases be methylated to different extents depending on the host specificity and that methylation is the chemical basis or at least an absolute requirement for host-controlled modification of phage T1 by bacterial host cells lysogenic for P1 (Klein, 1965; Klein and Sauerbier, 1965).

While the first methyltransferase to be purified and characterised was a 5-methylcytosine methylase from *Bacillus subtilis* (Oda and Marmur, 1966), and the first detection of in vitro restriction reported used a crude extract from *Escherichia coli* cells harbouring the R factor system (Takano *et al.*, 1966), the validity of the hypothesis of host-controlled restriction and modification was finally established with the identification, isolation and characterisation of restriction endonucleases from the *Escherichia coli* strains K-12 (Meselson and Yuan, 1968) and B (Linn and Arber, 1968; Roulland-Dussoix and Boyer, 1969; Roulland-Dussoix and Boyer, 1969).

One should emphasise that the phenomenon of restriction and modification refers to the restriction of the propagation of phage and the host-dependent modification that accompanies it, and although the "classical" mechanism of restriction and modification of bacteriophages consist of endonucleolytic cleavage (restriction) of phage DNA when it is not methylated (modified) at certain sites, phenotypic restriction and modification of bacterial viruses can also occur by other means, *e.g.* by host dependent protein modifications that affect its adsorption abilities rather than reactions at the DNA level (Krüger *et al.*, 1980; Krüger *et al.*, 1980; Krüger and Schroeder, 1981).

## **1.2. Restriction / Modification Systems**

The interest in molecular mechanisms of DNA restriction and modification enzymes was in part driven by the hope that such enzymes would prove valuable as



reagents for the analysis of DNA molecules and as systems for the study of DNA-protein interactions. But while the *EcoK* and *EcoB* enzymes were among the first proteins identified capable of specific recognition of DNA sequences, they did not display a simple relationship between the location of their recognition sites and the site of DNA cleavage as had been hoped for.

The additional purification and characterisation of the R factor restriction endonuclease (Yoshimori, 1971) and of a restriction endonuclease from *Haemophilus influenzae* (Kelly and Smith, 1970; Smith and Wilcox, 1970) however soon established a second class of restriction endonucleases which exhibited the simple mode of site specific catalysis initially anticipated for the *EcoKI* and *EcoBI* enzymes, and consequently produced discrete fragments of DNA.

While *EcoKI* and *EcoBI* were found to have an absolute requirement for ATP and  $Mg^{2+}$  in the restriction reaction, these new enzymes further distinguished themselves by only requiring  $Mg^{2+}$  as cofactor. Consequently the R/M systems were subdivided in ATP-requiring, type I, and ATP-independent, type II enzymes (Boyer, 1971).

This simple classification of restriction-modification systems into type I and type II enzymes according to cofactor requirements and enzymatic behaviour first had to be revised when it became apparent that the ATP-requiring enzymes could be subdivided into ATP-hydrolysing type I and non-hydrolysing type III enzymes (Arber, 1970; Haberman, 1974; Kauc and Piekarowicz, 1978). Although the reason for this subdivision has been thrown into doubt by recent results (Meisel *et al.*, 1995; Saha and Rao, 1995), the reclassification of *EcoP1* and others as type III enzymes is still justified by differences in genetic organisation, subunit composition, site specificity and enzymatic mechanism.

The usefulness of type II restriction endonucleases in the analysis of DNA and in the construction of recombinant DNA molecules sparked interest into the search for new enzymes with new specificities, and led to the discovery of more than 2700 new restriction-modification systems with more than 200 different specificities, the vast majority of which are type II enzymes (Roberts and Macelis, 1996). It should be emphasised however that the assays used to discover new R/M systems are designed to pick up type II enzymes. Although this is commercially sensible, this bias makes it impossible to draw any conclusions about the distribution and frequency of the different R/M systems in nature from the number of systems discovered.



The search for type II endonucleases with new specificities also led to the discovery of restriction-modification systems which do not seem to fit into any of the three established classes. The ones resembling type II enzymes in their cofactor requirement and simple mode of action (*i.e.* consisting of an endonuclease and a methylase acting independently) have been grouped together, and since they cleave the DNA on one side of an asymmetric recognition site were classified as type IIs (shifted cleavage) enzymes.

Every classification however is necessarily an artificial division of the variety of systems found in nature - as illustrated by the biochemical characterisation of a group of restriction endonucleases which in addition to their target site need a second effector site for efficient cleavage. This group contains both type II and type IIs enzymes (among them *NaeI*, *NarI*, *BspMI*, *HpaII*, *SacII*, *EcoRII*, *AtuBI*, *Cfr9I*, *SauBMKI*, *Eco57I*, and *Ksp632I*) and suggestions have been made to classify them as type IIE enzymes (Jo and Topal, 1995). However the restriction endonuclease *Eco57I* has also been found to be both stimulated a 100-fold by the presence of low amounts of AdoMet and to contain methylase activity (Petrusyte *et al.*, 1988), resembling type III enzymes in this respect but without the requirement for ATP, and it was therefore suggested to classify it as type IV (Janulaitis *et al.*, 1992).

Other examples of unusual restriction endonucleases include *Bsp24I* (stimulated by AdoMet, possible methylase activity, cleavage on both sides outside of its recognition sequence (Degtyarev *et al.*, 1993)) and *BcgI* (complex of the form A<sub>2</sub>B, requires AdoMet, cleavage on both sides outside of its asymmetric and interrupted recognition sequence 5'-CGA N<sub>6</sub> TGC-3' (Kong *et al.*, 1993; Kong *et al.*, 1994)).

Given the apparent variety of R/M systems and the occurrence of "intermediate forms" between the standard classes of R/M systems, it might be more fruitful to think of these enzymes as modular systems, where functions like DNA binding, cleavage, methylation, ATPase activity and others might reside in separate domains, and could be combined in new ways. This has already been shown in some respects for type I systems (Fuller-Pace *et al.*, 1984; Fuller-Pace and Murray, 1986; Gann *et al.*, 1987; Gubler *et al.*, 1992), for the type II methyltransferases *M.BspRI* and *M.BsuRI* (Kim *et al.*, 1991), and for the type IIs endonucleases *LlaI* (Hill *et al.*, 1991) and *FokI* (Kim and Chandrasegaran, 1994; Kim *et al.*, 1996).



### 1.3. Classification of Restriction / Modification Systems

#### 1.3.1. according to complexity and cofactor requirements

This classical way of organising a large number of R/M-systems into four different classes (summarised in Fig. 1a and Table 1a) clearly shows how the increasing complexity of the R/M-systems is accompanied by the separation of function onto different subunits. This and the existence of intermediate forms also hint at a possible pathway for the divergence of R/M-systems from a common ancestor during evolution to account for the different types of systems observed.

##### 1.3.1.1. Type II

Type II are the simplest of the three types of R/M-systems. They comprise two enzymes, an endonuclease encoded by the *res* gene, and a separate methylase, the product of the *mod* gene. The *res* and *mod* genes are often plasmid-borne and may be in either order relative to one another, transcribed convergently, divergently, or in the same direction by two promoters or one.

The only cofactor required for restriction is  $Mg^{2+}$ , and the endonuclease recognises a palindromic sequence of four to eight base pairs (which may be continuous or interrupted) and cleaves the DNA symmetrically within the sequence. The endonuclease generally functions as a homodimer, an association that facilitates the co-ordinated cleavage of both strands.

The corresponding Type II methyltransferases act generally as monomers (although dimers have been reported for the *DpnII* system (de la Campa *et al.*, 1987)), use AdoMet as methyl-donor for modification, and methylate the duplex one strand at a time, consistent with the hemimethylated nature of their customary substrates.

##### 1.3.1.2. Type IIs

Type IIs R/M-Systems (shifted cleavage) resemble Type II systems in that the endonuclease and methyltransferase act independently and have similar cofactor requirements, but otherwise are quite different (for review see Szybalski *et al.*, 1991). They recognise an asymmetric four to seven base pairs wide sequence, and the



asymmetry of their target has consequences for both the restriction and the modification component of these systems.

The endonucleases are about twice the size of type II endonucleases, act as monomers, and cleave on only one site of their target at a defined distance - under 20 bp - from the sequence.

Modification is usually accomplished by a pair of methyltransferases (in the case of *FokI* the methyltransferases are fused together in a single, bifunctional enzyme), one for each strand, since the asymmetry of their target enables them to bind only in one orientation, and in some systems different nucleotides are methylated on each strand, *e.g.* an A on one strand and a C on the other.

#### 1.3.1.3. Type III

Only four Type III systems have been isolated (for a review see Bickle, 1993), and although they comprise two genes *res* and *mod* like the type II enzymes, they are very different systems since these genes do not encode independent enzymes. The *mod* gene encodes an active methylase which combines with the product of the *res* gene to form a heterodimer with both methyltransferase and restriction activity (Bickle, 1987). The M subunit therefore confers sequence specificity for both restriction and modification. Methylation requires AdoMet, while cleavage requires  $Mg^{2+}$  and ATP and is stimulated by AdoMet (Yuan and Reiser, 1978).

The recognition sequences are asymmetric, uninterrupted, and five to six nucleotides in length. Cleavage occurs approximately 25 nts to one side of the sequence. Only one strand of the recognition sequence becomes methylated, in apparent violation of the rule that both strands must be methylated to preserve modification during replication. However, cleavage takes place only when two unmodified sites are present in the DNA, in opposite directions. Since one site or another remains modified after passage of the replication fork, modification is preserved during replication (Krüger *et al.*, 1990; Meisel *et al.*, 1992).

#### 1.3.1.4. Type I

Type I systems have been found in *Escherichia coli* and its relatives, *Citrobacter* and *Salmonella*, and more recently in *Mycoplasma* (Dybvig and Yu, 1994), *Klebsiella* (Rutebuka *et al.*, 1995; Valinluck *et al.*, 1995), *Bacillus* (Xu *et al.*, 1995), *Pasteurella* (Highlander and Garza, 1996), *Haemophilus* (Fleischmann *et al.*, 1995) and possibly



even in the thermophile *Methanococcus*, where an ORF with 61.4 % sequence similarity to *EcoR124II* has been identified (Bult *et al.*, 1996). Currently, Type I systems are grouped into four families on the basis of complementation tests, immunological cross-reactivity and DNA-hybridization of their coding sequences, and a recent screen of natural isolates of *Escherichia coli* with family specific DNA probes for the type IA, IB, and ID families indicated that approximately 45% of them carry systems belonging to these families (Barcus *et al.*, 1995).

Type I enzymes show an even higher degree of the separation of functions than Type III enzymes - they are large multisubunit enzymes where restriction, modification and sequence specificity functions are encoded by different genes (*hsdR*, *hsdM*, and *hsdS* respectively), as had been shown early on by complementation analysis using different mutants of the *EcoKI* and *EcoBI* systems (Boyer and Roulland-Dussoix, 1969; Glover and Colson, 1969; Glover, 1970).

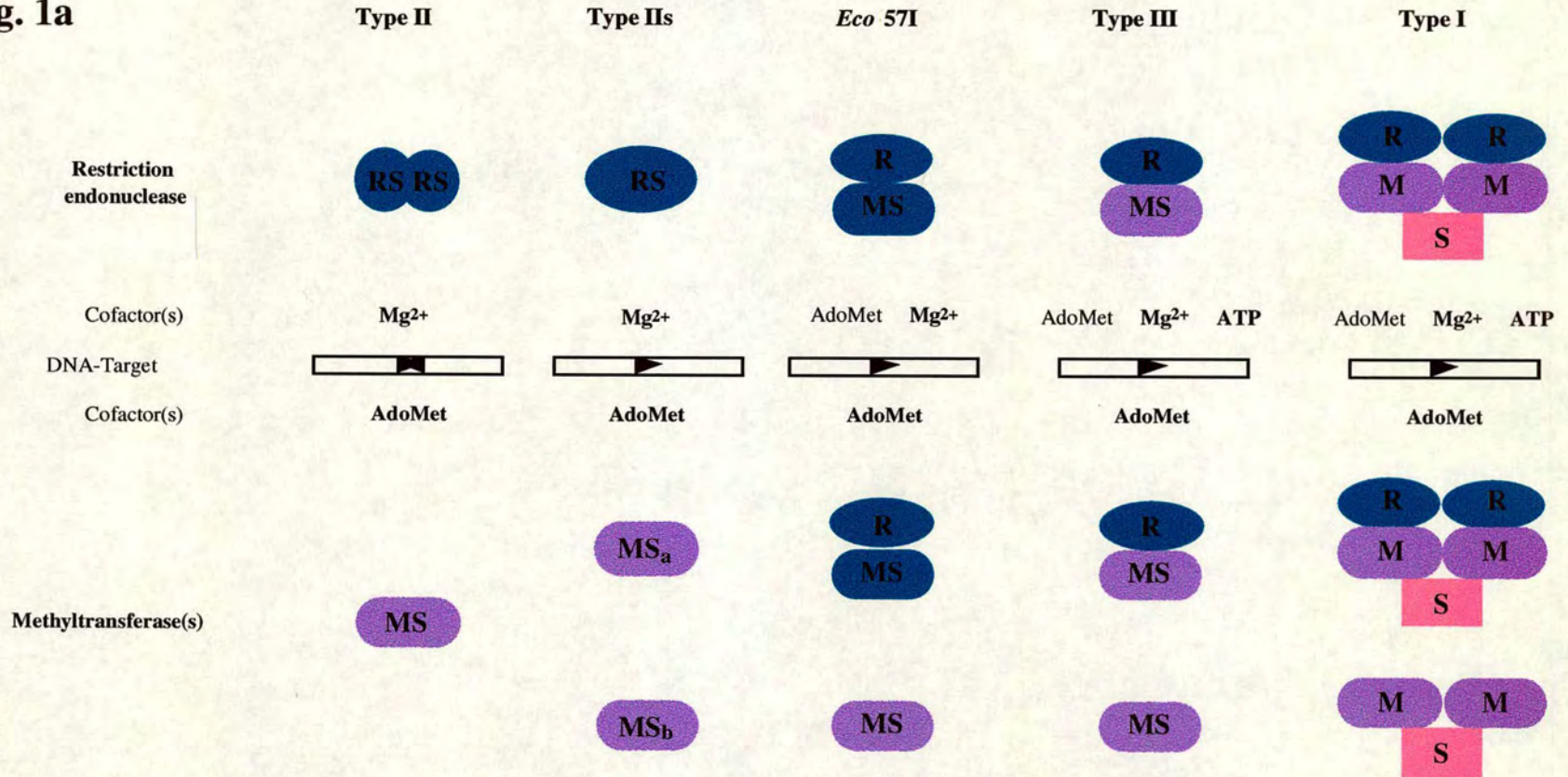
These systems recognise asymmetric sequences comprising one trinucleotide and one tetra- or pentanucleotide compartment separated by a non-specific spacer of 6 to 8 nt, and the methylated bases are 10 or 11 nt (roughly one turn of the DNA helix) apart. The recently characterised *StySKI* system (Thorpe *et al.*, 1997) is unusual in that its recognition sequence consists of two tetranucleotide compartments.

In the absence of the HsdR subunit, the HsdM and HsdS subunits form an ATP-independent methyltransferase of the stoichiometry  $M_2S_1$ , which only requires AdoMet as cofactor (Lautenberger and Linn, 1972; Suri and Bickle, 1985; Taylor *et al.*, 1992; Dryden *et al.*, 1993). Methylation activity resides in the HsdM subunit, which contains the AdoMet binding site, while sequence specificity is determined by the HsdS subunit.

In the presence of the HsdR subunit, an active complex with both restriction and modification activity is formed. Its stoichiometry is believed to be  $R_2M_2S_1$  for the best studied of the type I systems, *EcoKI* and *EcoBI* (Weiserova *et al.*, 1993; Dryden *et al.*, 1997), which require ATP, AdoMet and  $Mg^{2+}$  as cofactors in the restriction reaction (Linn and Arber, 1968; Meselson and Yuan, 1968; Roulland-Dussoix and Boyer, 1969; Roulland-Dussoix and Boyer, 1969). The  $R_1M_2S_1$  complex reported for *EcoR124I* (Janscak *et al.*, 1996) has been found to be inactive (unpublished observations). Restriction of unmodified DNA is achieved by a complex process involving the translocation of DNA and cleavage at a site typically several kilobases away from the target site, and is accompanied by extensive ATP hydrolysis.



**Fig. 1a**



**Fig. 1a: Classes of Restriction/Modification Systems according to complexity and cofactor requirements**

Different colours represent different proteins or subunits, while R, M, and S denote the restriction, modification, and specificity function respectively. The two different methyltransferases encoded by the Type IIs systems are indicated by indices. Cofactors are shown for both the restriction and methylation reaction: bold lettering indicates that the cofactor is essential for the respective reaction, while plain lettering indicates a stimulating effect only. The symmetry of the respective target recognition site is mirrored by the symmetry of the symbols used:

symmetrical target recognition site

asymmetrical target recognition site



**Table 1a: Characteristics of the four classical types of R/M-systems**

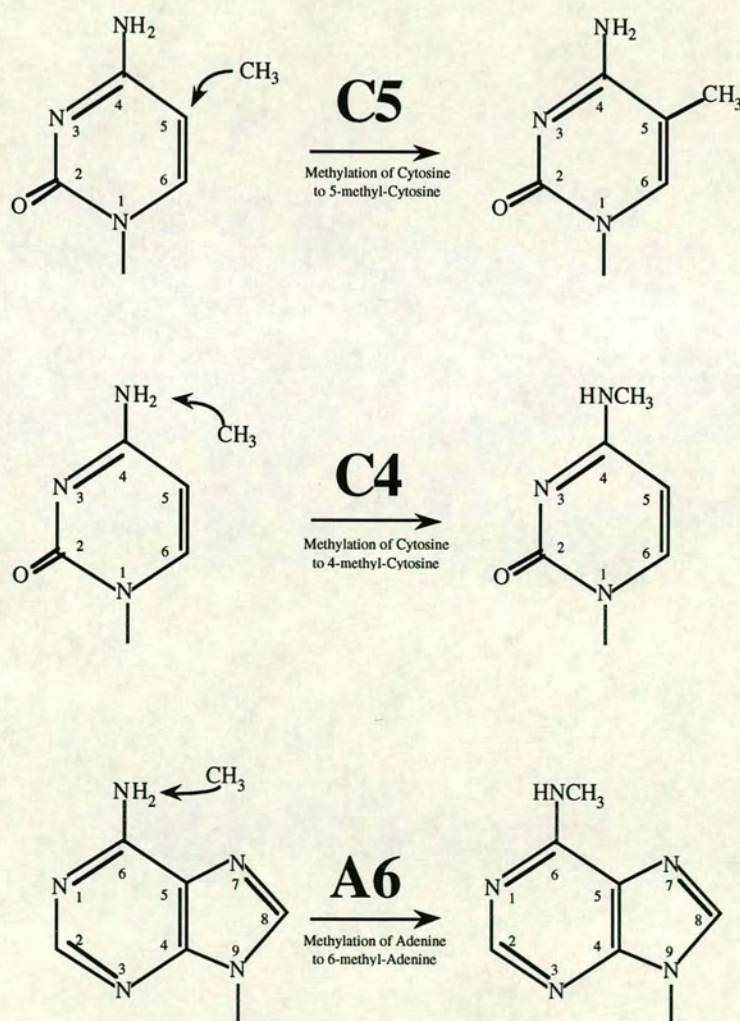
	Type II	Type IIs	Type III	Type I
<b>Restriction / Modification activities</b>	One restriction enzyme, one modification enzyme	One restriction enzyme, two modification enzymes	Single enzyme with R and M activity (as RM heterodimer) or M activity (as M Monomer)	Single enzyme with R and M activity (as R <sub>2</sub> M <sub>2</sub> S heteromer) or M activity (as M <sub>2</sub> S heteromer)
<b>Protein structure</b>				
a) Methyl-transferase	simple (monomer)	simple (monomer)	simple (M monomer) or complex (2 different subunits, RM heterodimer)	complex (R <sub>2</sub> M <sub>2</sub> S or M <sub>2</sub> S)
b) Restriction endonuclease	simple (homodimer)	simple (monomer, twice as large as Type II)	complex (RM heterodimer)	complex (R <sub>2</sub> M <sub>2</sub> S)
<b>Requirements for restriction</b>	Mg <sup>2+</sup>	Mg <sup>2+</sup>	Mg <sup>2+</sup> , ATP, 2 target sites in inverse orientation (stimulated by AdoMet)	Mg <sup>2+</sup> , ATP (stimulated by AdoMet)
<b>Requirements for methylation</b>				
a) Methyl-transferase	AdoMet	AdoMet	AdoMet	AdoMet
b) Restriction endonuclease	no methylation activity	no methylation activity	AdoMet (stimulated by Mg <sup>2+</sup> , ATP)	AdoMet (stimulated by Mg <sup>2+</sup> , ATP)
<b>Recognition sequence</b>	two-fold symmetry	asymmetric, uninterrupted	asymmetric, uninterrupted	asymmetric, interrupt
<b>Cleavage</b>	complete	-	incomplete	-
<b>Methylation site</b>	Recognition sequence	Recognition sequence	Recognition sequence	Recognition sequence
<b>Cleavage site</b>	within recognition site	< 20 bp in 3'-direction from recognition site	24-29 bp in 3'-direction from recognition site	up to several kb away from recognition site
<b>Methylation state of substrate</b>				
a) Methyl-transferase	un- or hemimethylated	un- or hemimethylated	un- or hemimethylated	un- or hemimethylated (hemimethylated for Type IA and IC)
b) Restriction endonuclease	unmethylated	unmethylated	unmethylated	unmethylated



### 1.3.2. according to the type of methylation

While the classification of R/M-systems according to complexity emphasises the differences between the systems and thereby mirrors the increasing divergence of R/M-systems during evolution, the classification according to the type of methylation puts emphasis on their similarities and a possible common origin for the R/M-systems.

Three types of methylation are found among the R/M-systems (see Fig. 1b): methylation of the exocyclic amino group of adenosine to N6-meA (Dunn and Smith, 1955), methylation of the exocyclic group of cytosine to N4-meC (Janulaitis *et al.*, 1983), and methylation of the 5-C ring carbon of cytosine to 5-meC (Wyatt, 1950).



**Fig. 1b: Types of Methylation**



It has been proposed that the use of N4-meC rather than 5-meC was an adaptation to thermophilia, which would avoid the hypermutability associated with 5-meC (Ehrlich *et al.*, 1985). Deamination of 5-meC is a highly temperature-dependent reaction which produces thymine, resulting in G=C to A=T transition mutations upon DNA replication, as has been observed even at mesophilic temperatures (Duncan and Miller, 1980; Wang *et al.*, 1982; Coulondre *et al.*, 1987). N4-meC (as well as C) also undergoes spontaneous deamination, but the product is uridine, which is subject to correction by uridine-specific DNA repair pathways (Duncan and Miller, 1980). An initial screening of thermophilic bacteria showed that about half of the strains tested contained N4-meC in their DNA (Ehrlich *et al.*, 1985), and N4-meC is also present in the DNA of many mesophilic species (Butkus *et al.*, 1985; Ehrlich *et al.*, 1987). The recent cloning of the whole genome of the thermophilic *Methanococcus jannaschii* however revealed several open reading frames with pronounced similarity to C5-methyltransferases.

Other types of methylation of DNA bases are known, *e.g.* the methylation of guanine to O6-methylguanine, but no R/M system with this type of methylation has been identified. The likely reason is that O6-methylation of guanine induces G to A transitions, and cells possess O6-methylguanine-DNA methyltransferases that repair DNA damaged in this way by alkylating agents (Takano *et al.*, 1991).

While all known type I and type III systems methylate the extracyclic amino group of adenosine to N6-methyl-adenosine, examples for all three types of methylation can be found in type II systems. The grouping of the DNA methyltransferases according to the type of methylation and comparisons of their deduced amino acid sequences within these groups as well as with other AdoMet binding proteins (Som *et al.*, 1987; Chandrasegaran and Smith, 1988; Posfai *et al.*, 1988; Klimasauskas *et al.*, 1989; Lauster *et al.*, 1989) not only led to the discovery of sequence motifs common to all methyltransferases (and therefore implicated in AdoMet binding), but also to the discovery of sequence motifs specific for these groups (and therefore implicated in the catalysis of this specific type of methylation), and established an alternative classification of the modification enzymes into C4-, C5- and A6-methyltransferases.

The best characterised methyltransferases are found in the C5-group, where a total of six highly conserved and four lesser conserved blocks of sequence can be identified that are all arranged in a common order (Fig. 1c; for a review see Kumar *et al.*, 1994). The structures of these methyltransferases consist of a variable-length



N-terminal arm followed by eight well-conserved blocks each separated by small variable-length regions. A large variable-length segment of 90 to 270 amino acids then follows, which was shown by studies on multispecific, phage methyltransferases (Trautner *et al.*, 1988; Trautner *et al.*, 1988) and by experiments involving the swapping of this variable region between enzymes with different specificities (Klimasauskas *et al.*, 1991) to contain the target-recognising domains (TRDs) of these enzymes. After this are two more conserved blocks, and a variable-length C-terminal segment completes the sequence .



The 10 conserved motifs in *HhaI* methyltransferase (x denotes any amino acid)

**Fig. 1c**

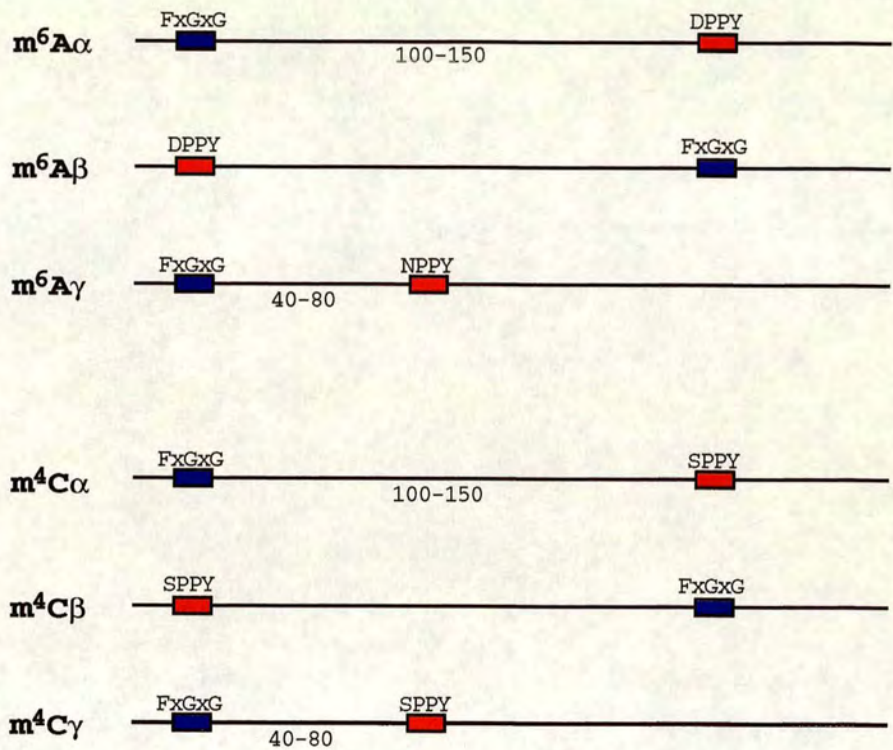
Motif I is common among AdoMet binding proteins and was therefore inferred to be part of the AdoMet binding site. The strongest conservation however is found in motif IV which contains an invariant dipeptide proline-cysteine, which suggested that this motif is involved in the transfer of a methyl group from AdoMet onto cytosine (Lauster, 1989).

C4-methyltransferases showed a surprising lack of similarity to their C5 counterparts, and proved to be more similar to A6-methyltransferases. Sequence comparisons using A6- and C4-methyltransferases revealed two highly conserved motifs, one similar to motif I, the other similar to motif IV in C5 methyltransferases (Chandrasegaran and Smith, 1988; Klimasauskas *et al.*, 1989; Lauster, 1989). Contrary to C5-methyltransferases the relative position of these motifs to each other, with motif I close to the N-terminus and motif II about halfway towards the C-terminus, is not fixed.

An extensive comparison of the deduced amino acid sequences with consideration of the relative position of the highly conserved motifs led to the subgrouping of the A6- and C4-methyltransferases into three groups  $\alpha$ ,  $\beta$ , and  $\gamma$  (see Fig. 1d; Wilson, 1992). The  $\gamma$ -group is most similar to the C5-methyltransferases, *e.g.* the variable region implicated in DNA-binding is located towards the C-terminus, while the variable region has a central location in the  $\alpha$ - and  $\beta$ -group. The implication



of this is that the proposed "methylation domain" consists either of two domains or of two half-domains.



A6- and C4-methyltransferases are separated into three groups  $\alpha$ ,  $\beta$ , and  $\gamma$  based on the relative location of motif I (blue) and motif II (red) to each other in their primary sequence.

**Fig. 1d**

More exhaustive sequence alignments (this thesis; Noyer-Weidner *et al.*, 1994; Dryden *et al.*, 1995) became possible due to the increasing number of methylase genes sequenced, and led to the identification of additional motifs and highly conserved amino acids among C4- and A6-methyltransferases.

There are some significant irregularities in the distribution of methyltransferases onto the different classes of R/M systems. While Type II methyltransferases comprise C5-, A6-, and C4-methyltransferases, all Type I and Type III methylases characterised so far are A6-methyltransferases. To date, no C4-methyltransferase of the  $\gamma$ -group has been found, although this may be due to the relatively small number of C4-methyltransferases known.



To summarise, the classification of the methyltransferases according to their mode of methylation strongly correlates with their deduced primary sequence and allows the definition of conserved motifs implicated in binding of the cofactor AdoMet and in catalysis. The arrangement of these conserved motifs and the location of the variable regions in turn lead to the definition of separate domains for methylation and target recognition. Domain swapping experiments confirmed these deductions.

While the methyltransferases show extensive similarities at the primary sequence level, the corresponding restriction endonucleases show little significant sequence similarity with either their cognate methyltransferase nor with other endonucleases recognising the same sequence (Chandrasegaran and Smith, 1988). This was interpreted as indicative of a common origin for the methyltransferases, while the endonucleases were thought to have arisen independently. A recent more refined analysis of the available sequence data however showed that the endonucleases are more similar to each other than would be expected by chance (Jeltsch *et al.*, 1995), and the lack of sequence similarity with their cognate methyltransferase might be indicative of the different ways in which they interact with their DNA targets, *e.g.* the endonucleases as dimers, and the methyltransferases as monomers.

## **1.4. Mechanism of Restriction and Modification**

### **1.4.1. Type II enzymes**

Most of our knowledge about the mechanisms of restriction and modification is derived from studies on Type II enzymes. Despite their simplicity in regard to complexity and cofactor requirements, Type II enzymes are not simple enzymes - endonucleases have to recognise their target sequences with high specificity among a vast excess of dissimilar, similar, and identical but modified sequences, while the corresponding methyltransferases have to recognise the same target sequence and in the case of C5-methyltransferases have to methylate a base that is seemingly inaccessibly buried in the DNA helix.

The wealth of biochemical, mutational, and kinetic data available for Type II enzymes has in recent years been complemented by the determination of the crystal structures of some of the enzymes. So far the crystal structures of the Type II



endonucleases *R.EcoRI* (Frederick *et al.*, 1984; McClarin *et al.*, 1986; Kim *et al.*, 1990), *R.EcoRV* (Winkler *et al.*, 1993), *R.BamHI* (Newman *et al.*, 1994), *R.PvuII* (Athanasiadis *et al.*, 1994; Cheng *et al.*, 1994), and *R.Cfr10I* (Bozic *et al.*, 1996), of the Type IIs endonuclease *R.FokI* (Kita *et al.*, 1989; Wah *et al.*, 1997), and of the Type II methyltransferases *M.HhaI* (Cheng *et al.*, 1993), *M.TaqI* (Schluckebier *et al.*, 1995c), *M.HaeIII* (Reinisch *et al.*, 1995; Reinisch, 1996), and *M.PvuII* (Gong *et al.*, 1997) have been elucidated, some of them both in the presence and absence of substrate and cofactor (see table 1b).

**Table 1b:**

enzyme	recognition sequence	structure - DNA	structure + DNA
<i>R.EcoRI</i>	G↓AATTC	√	√
<i>R.EcoRV</i>	GAT↓ATC	√	√
<i>R.BamHI</i>	G↓GATCC	√	√
<i>R.PvuII</i>	CAG↓CTG	√	√
<i>R.Cfr10I</i>	R↓CCGGY	√	-
<i>R.FokI</i>	GGATG (9↓13)	-	√
<i>M.HhaI</i>	G <u>C</u> GC	√	√
<i>M.TaqI</i>	TCG <u>A</u>	√	-
<i>M.HaeIII</i>	GG <u>C</u> C	-	√
<i>M.PvuII</i>	CAG <u>C</u> TG	√	-

↓ indicates the location of the scissile bond,  
while the underscore marks the base that is methylated.

#### 1.4.1.1. Endonucleases

Endonucleases, like many other proteins interacting with specific DNA sequences, seem to locate their target sites on the macromolecular substrate via a facilitated diffusion mechanism in which they first bind non-specifically and then slide along the DNA until they encounter their cognate target, thereby accelerating their association rates beyond their three-dimensional diffusion limit (Ehbrecht *et al.*, 1985; Maass *et al.*, 1986; Berkhout and van Wamel, 1996). Appropriately, studies on



*R.EcoRI* suggest that the endonuclease can bind to DNA in two modes: one loose and non-specific, allowing for linear diffusion, the other one tight, specific, and immobile, leading to DNA cleavage (Jeltsch *et al.*, 1994). Conforming with this view crystal structures of *R.EcoRV* without DNA and in complex with both cognate and non-cognate DNA show that considerable conformational changes occur in both protein and DNA on complex formation (Winkler *et al.*, 1993). Non-cognate DNA is bound in an essentially B-form conformation with protein-DNA interactions that are prototypic for non-specific DNA binding, while bound cognate DNA deviates considerably from canonical B-form DNA and specific protein-DNA contacts are made. However DNA distortion is not a prerequisite for specific complex formation, as is evident from crystal structures of *R.PvuII* and *R.BamHI* in complex with their cognate DNA.

In their functional dimeric form, the endonucleases adopt a U-shaped conformation with a prominent cleft for DNA binding. An early hypothesis relating the position of the scissile bonds with the way a restriction enzyme approaches the DNA was put forward by Anderson (Anderson, 1993). He noticed that for 5' overhangs the scissile bonds are on the major-groove side of the double helix, while they are on the minor-groove side for 3' overhangs; for blunt ends they are nearer the minor groove, but practically on opposite sides of the double helix. Therefore enzymes which leave 5' overhangs would approach the DNA from the major-groove side and those that leave 3' overhangs would approach it from the minor-groove side. Enzymes that leave blunt ends would approach from one side or the other depending on the intrinsic conformation of the unbound side. The structures of endonucleases bound to DNA published so far support this hypothesis (although no enzyme leaving 3' overhangs has been crystallised so far).

The crystal structures show that despite the lack of any significant primary sequence similarity, the two blunt-cutting enzymes *R.PvuII* and *R.EcoRV* resemble each other, as do the two enzymes leaving 5'-overhangs *R.BamHI* and *R.EcoRI*. *R.Cfr10I*, which recognises a degenerate target and cleaves DNA to produce 5'-overhanging ends, resembles *R.EcoRI* but also exhibits some structural similarities to *R.EcoRV*.

Although no distinct domain structure is apparent, comparative studies of the endonuclease structures show a striking degree of modularity where different functions can be attributed to different regions in the structures, which have



accordingly been described as the subunit interface region or dimerisation subdomain, the DNA sequence recognition region, and the catalytic region (Cheng *et al.*, 1994).

This modularity is most apparent in the determination of sequence specificity. Although the DNA-recognition elements are surprisingly diverse, among the enzymes leaving 5'-overhanging ends the inner four base-pairs within a hexanucleotide sequence are recognised by one structural element ("central recognition module", equivalent to the extended chain and inner recognition helix in *R.EcoRI*), while the outer base-pairs are recognised by another element ("outer recognition module", equivalent to the outer recognition helix in *R.EcoRI*). In *R.Cfr10I*, which recognises the degenerate sequence RCCGGY, the presence of a structurally equivalent "outer recognition module" is not evident. *R.EcoRV* and *R.PvuII*, which leave blunt ends after restriction, use a similar modular organisation of DNA-recognising elements: the central two base-pairs are recognised by one element (although they are not directly recognised in *R.EcoRV*) and the outer four base-pairs by another (Bozic *et al.*, 1996).

Comparisons between the different available structures of *R.EcoRV* reveal that the relative position and orientation of the two DNA binding subdomains are highly variable, due to the existence of flexible segments linking the dimerisation and the DNA binding subdomains with each other. This structural flexibility may allow the enzyme to translocate from one non-cognate binding site to the next by moving one of its two DNA binding subdomains at a time. The flexible DNA recognition loop can probe the local DNA sequence during this process, while contact with the target sequence would lead to a transition to the cognate binding mode (Winkler *et al.*, 1993).

All the endonucleases share a common central core structure consisting of a mixed five-stranded  $\beta$ -sheet, with the active site located at one end, flanked by helices (Cheng *et al.*, 1994; Newman *et al.*, 1994; Aggarwal, 1995). Comparisons of the active sites of the endonucleases reveals a conserved structural motif (E/D) $X_n$ (E/D)ZK (where  $n$  varies from 9 to 18 residues,  $X$  is any residue and  $Z$  is a hydrophobic one) of one basic and two acidic groups close to the reactive phosphodiester group, where the conserved acidic pair is well positioned to co-ordinate the catalytic cofactor  $Mg^{2+}$  (Anderson, 1993; Winkler *et al.*, 1993; Aggarwal, 1995).

The cleavage reaction is thought to proceed as an in-line  $S_N2$  attack on the scissile phosphate by an activated water molecule. Activation of the water molecule was suggested to occur either by  $Mg^{2+}$  and one of the acidic residues (Selent *et al.*, 1992), or by the phosphate group adjacent to the scissile bond in a substrate-assisted



reaction (Jeltsch *et al.*, 1992). Experiments with oligonucleotides in which this phosphate was substituted by a methyl phosphonate support the idea of substrate-assisted catalysis in the cleavage reaction for several of the endonucleases tested, but not for all of them (Jeltsch *et al.*, 1995). The essential cofactor  $Mg^{2+}$  is also thought to polarise the P-O bond (thereby increasing the electrophilicity of the phosphorus atom), to stabilise the transition state and to supply the water molecule for protonation of the leaving group (Grasby and Connolly, 1992; Selent *et al.*, 1992; Kostrewa and Winkler, 1995; Pingoud *et al.*, 1995).

Since the active site residues of enzymes generating blunt or 5'-overhanging ends, respectively, are similarly arranged within topologically equivalent secondary structure elements, different dimerisation modes are used to achieve the correct positioning of the active sites. *R.EcoRI*, *R.BamHI* and *R.Cfr10I* use similar dimer interfaces to separate the active sites by about 18 Å (4 bp), while *R.EcoRV* and *R.PvuII* possess different dimerisation subdomains with no similarity to each other (Bozic *et al.*, 1996).

As a consequence of the structural information available for Type II restriction endonucleases it is becoming increasingly clear that restriction enzymes, which in general do not share significant sequence identity, have structural and mechanistic features in common. Indeed, using a genotypic and phenotypic analysis it has been possible to suggest a common evolutionary history for these enzymes (Pingoud *et al.*, 1995). The crystal structures of endonucleases in the presence and absence of substrate DNA (both cognate and non-cognate) reveal that a wide variety of conformational changes can occur. The DNA might be kinked and underwound (*R.EcoRI*), bent (*R.EcoRV*), or bound in a largely undisturbed B-form (*R.PvuII*). Similarly, the enzyme might undergo extensive conformational changes including rotation of subunits and folding of disordered regions as in the case of *R.BamHI*, where the most striking conformational change is the unravelling of carboxyl-terminal alpha helices to form partially disordered "arms". The arm from one subunit fits into the minor groove while the arm from the symmetry-related subunit follows the DNA sugar-phosphate backbone (Newman *et al.*, 1995). The complexity of these "simple" enzymes is increased by the fact that some endonucleases require one  $Mg^{2+}$  ion at the active site for phosphodiester bond hydrolysis, while others require two  $Mg^{2+}$  ions (Baldwin *et al.*, 1995; Kostrewa and Winkler, 1995; Vipond *et al.*, 1995), or that a high affinity  $Mg^{2+}$  binding site is formed only with the specific DNA (*R.EcoRV*).



Others again require the presence of effector sites (Type IIe enzymes), indicating additional DNA binding domains and subunit interactions.

#### 1.4.1.2. Methyltransferases

Although much of the early work has been done on Type I methyltransferases, the best characterised methyltransferases to date are Type II methyltransferases. The wealth of data available due to sequence alignments and extensive biochemical and mutational characterisation, and its combination with structural information derived from crystal structures offers a uniquely detailed picture of how Type II methyltransferases operate. The similarities of Type II methyltransferases with other DNA, RNA, protein and small molecule methyltransferases indicate that many AdoMet-dependent methyltransferases have a common catalytic domain structure (Schluckebier *et al.*, 1995a), and make tertiary structure predictions from sequence alignments possible.

#### C5-Methyltransferases

One of the best characterised methyltransferases among the Type II enzymes is *M.HhaI*, a C5-methyltransferase from *Haemophilus haemolyticus*. After the discovery of the *HhaI* R/M-system in 1976 (Roberts *et al.*, 1976), its subsequent purification and biochemical characterisation (Mann and Smith, 1979; Wu and Santi, 1987; Wu and Santi, 1988) established that *M.HhaI* methylates the central cytosine of its recognition sequence 5'-GCGC-3' in an ordered Bi-Bi reaction. The reaction starts with the binding of *M.HhaI* to DNA followed by binding of the methyl donor AdoMet. After the essentially irreversible transfer of the methyl group onto the substrate DNA, the product AdoHcy leaves the ternary complex first, followed by dissociation of the enzyme from the methylated DNA. AdoMet and AdoHcy were found not to bind to the free enzyme, and only poorly to a complex of *M.HhaI* with methylated DNA.

Somewhat contradictory to this, the purification of *M.HhaI* by a novel method in 1992 using an overexpressing *Escherichia coli* strain (Ho *et al.*, 1991) resulted in the isolation of two forms of the enzyme: one a form of *M.HhaI* capable of binding AdoMet, the other a stable and methylation-competent binary complex of *M.HhaI* and AdoMet. Instead of purifying *M.HhaI* from the soluble fraction and discarding the insoluble form of the methyltransferase, this novel purification method involved extraction of the methyltransferase from the disrupted cellular debris via a high-salt



step, while the  $\approx 30\%$  of *M.HhaI* present in soluble form in the supernatant was discarded.

These two seemingly contradictory results can be reconciled by proposing that *M.HhaI* exists in at least two states within the cell. The first one occurs free in the cytoplasm and is unable to bind AdoMet. Characterisation of this state would lead to the observed Bi-Bi mechanism. After (possibly unspecific) binding to DNA the methyltransferase would undergo a transition into a state capable of binding AdoMet. Purification and characterisation of this state would yield a form of *M.HhaI* either with AdoMet already bound or without bound AdoMet but capable of binding it, even in the artificially induced absence of DNA (see Kumar *et al.*, 1992, who report the isolation of two such forms). Addition of DNA would lead to rebinding, and the methylation reaction can proceed.

The existence of different states has already been shown for other methyltransferases. As early as 1975 an allosteric transition to an activated form upon binding of AdoMet had been shown for the Type I methyltransferase *EcoKI* (Yuan *et al.*, 1975). Circular dichroism spectroscopy had been used to demonstrate the existence of at least six ligand-induced conformational states in the Type IIs C5-methyltransferase *M.HgaI-2* (Baldwin *et al.*, 1994).  $^3\text{H}$  NMR studies on the *Escherichia coli* dam methyltransferase showed fit-induced rearrangements of the protein and the formation of an active complex capable of methylating DNA (Bergerat and Guschlbauer, 1990; Bergerat *et al.*, 1991), while spectroscopic studies on *M.EcoRI* implicate conformational changes upon DNA binding (Maegley *et al.*, 1992). There is also increasing evidence that methyltransferases first bind DNA non-specifically and then locate their target via a facilitated diffusion mechanism before they form a specific complex (Renbaum and Razin, 1992; Surby and Reich, 1992; Renbaum and Razin, 1995; Surby and Reich, 1996).

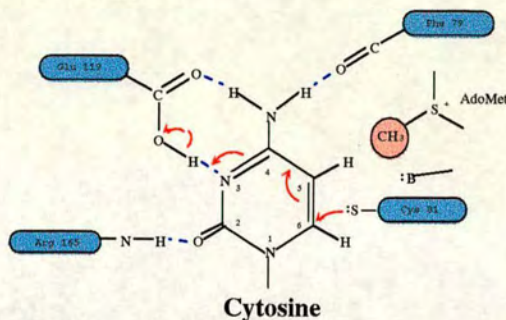
In analysing the reaction mechanism of the methyltransferase *M.HhaI*, Wu and Santi proposed a pathway for C5-cytosine methylation that proceeds through the formation of a transient, covalent protein-DNA complex (see fig. 1e; Wu *et al.*, 1985; Wu and Santi, 1987). A key feature of the proposed mechanism involves nucleophilic attack on the carbon-6 of the target cytosine by the thiol group of a cysteine. Direct support for their proposal came from the discovery that the cysteine of the invariant PC (motif IV) present in all C5-methyltransferases becomes covalently attached to DNA in trapped intermediates formed by *M.HaeIII* (Chen *et al.*, 1993), *M.EcoRII*



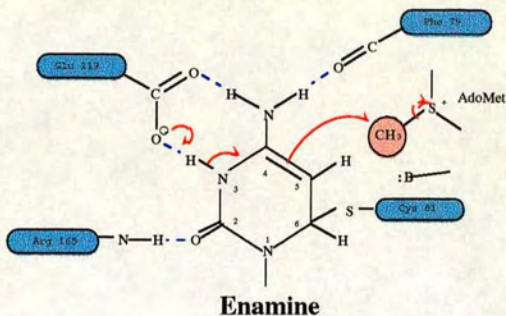
### Fig. 1e: Reaction pathway for C5-cytosine methyltransferases

based on the mechanism proposed by Wu and Santi (1985, 1987), modified by Baker *et al.* (1988) and Chen *et al.* (1993), and confirmed by the crystallographic work on *M.HhaI* (Cheng *et al.* 1993, Klimasauskas *et al.* 1994). The residues are inferred from the *M.HhaI* structure. A base (B:) that abstracts a proton from carbon-5 is needed for the  $\beta$ -elimination step. A water molecule bound to the side chains of Asn304 and Gln82 may serve as the required general base catalyst.

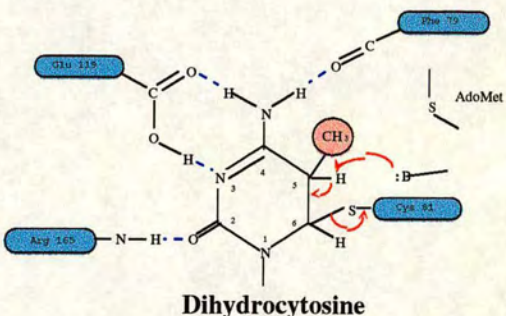
- a) Nucleophilic attack on target cytosine by the active site residue cysteine (Cys81 in *M.HhaI*).



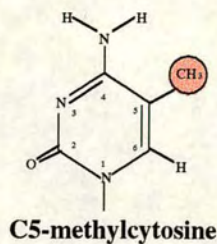
- b) Methyl-group acceptance by initial covalent intermediate (proton transfer from an acidic side residue (Glu119 in *M.HhaI*) avoids the formation of a high-energy carbanion intermediate).



- c) Abstraction of the proton at C5 again yields an enamine which undergoes  $\beta$ -elimination.



- d) The products C5-methylcytosine and S-adenosyl-L-homocysteine are released.





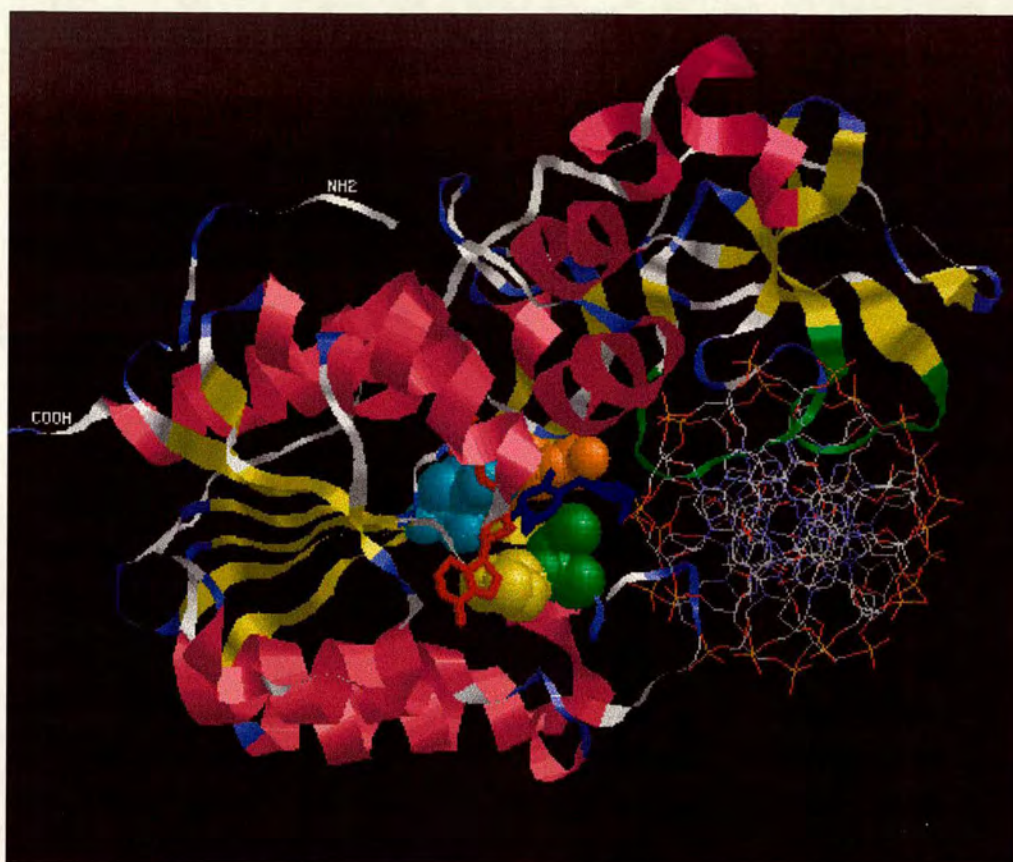
(Friedman and Ansari, 1992), human methyltransferase (Smith *et al.*, 1992), and *dcm* (Hanck *et al.*, 1993) in reactions with the suicide substrate 5-fluoro-2'-deoxycytosine. The stereochemistry of the methylation reaction however dictates that the attack on the 5-position has to occur at right angles to the plane of the pyrimidine ring, which is clearly not feasible with B-form DNA (Erlanson *et al.*, 1993).

The elucidation of the crystal structure of the C5-methyltransferase *M.HhaI* (Cheng *et al.*, 1993a; Cheng *et al.*, 1993b) in complex with DNA, however, brought a surprising answer to the problem of how to methylate a base that seems inaccessibly buried in the DNA helix: the target cytosine base is completely flipped out of the DNA duplex, and in its place the DNA is infiltrated by a glutamine and a serine side chain from the major and minor groove side respectively (Klimasauskas *et al.*, 1994; Patel, 1994; Phillips, 1994; Suck, 1994; Verdine, 1994; Winkler, 1994).

The crystal structure (see fig. 1f) shows that the polypeptide chain is folded into two domains, arranged in the form of a "C" with a wide cleft suitable to accommodate the DNA substrate. The larger catalytic domain contains two structurally similar nucleotide-binding pockets, one for AdoMet, the other for the target nucleotide (Malone *et al.*, 1995), while the smaller domain is responsible for both the sequence-specific recognition of DNA and the penetration of the DNA double helix to flip the target cytosine. This is in good agreement with results from earlier hybrid swap experiments with monospecific methyltransferases. These had previously established that the variable region in the primary sequence which encodes the smaller domain not only determines the sequence-specificity, but also the specific base to be methylated within the target sequence (Klimasauskas *et al.*, 1991; Mi and Roberts, 1992).

Of the ten conserved motifs observed in C5-methyltransferases, nine are contained in the catalytic domain. This domain has an  $\alpha/\beta$  structure, where the helices surround a mixed central  $\beta$ -sheet consisting of five parallel and two anti-parallel  $\beta$ -strands, and can be divided into two halves. The first half contains the conserved motifs I-III and X, bears a strong resemblance to the Rossmann mononucleotide-binding fold, and is mainly responsible for binding the cofactor AdoMet, while the second half contains the conserved motifs IV-VIII and is primarily involved in catalysis. As would be expected for  $\alpha/\beta$  structures, the majority of the invariant or highly conserved residues are located in loops at the carboxyl end of the  $\beta$ -strands, and as a consequence are clustered close to the active centre on the inner surface of the cleft.





**Fig. 1f:** *Crystal structure of M.HhaI in complex with DNA and SAH*

Molecule	Part	Display	Colour
<b>M.HhaI</b>	$\alpha$ -Helices	ribbon	magenta
	$\beta$ -strands	ribbon	yellow
	turns	ribbon	pale blue
	other residues	ribbon	white
	sequence recognition loops	ribbon	green
	<b>active site:</b>		
	Pro80	spacefill	yellow
	Cys81	spacefill	green
<b>DNA</b>	double helix	wireframe	CPK
	target cytosine	sticks	blue
<b>cofactor</b>	AdoHcy	sticks	red



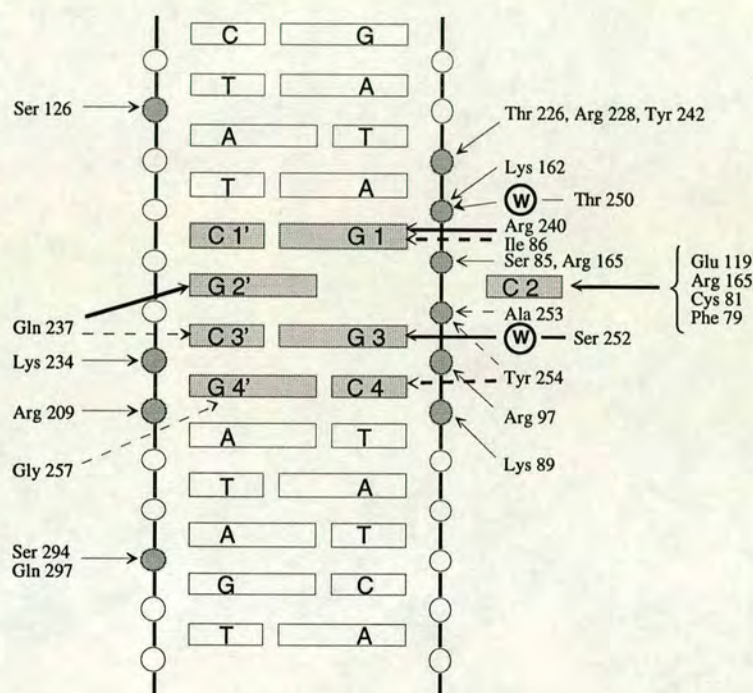
At the surface of the cleft, a pocket embedded in the catalytic domain next to the active site contains the AdoMet binding site. Binding of the cofactor AdoMet involves insertion of the methionine moiety into the pocket, while the adenosyl moiety is flanked by three conserved hydrophobic (Phe18 of motif I, Trp41 of motif II, and Leu100 of motif V) and two invariant negatively-charged side-chains (Glu40 of motif II and Asp60 of motif III). Of special interest is the water-mediated interaction between Glu119 and AdoMet observed in the ternary complex structures, which might explain why the binding of DNA by methyltransferases is generally more stable in the presence of AdoMet or AdoHcy or the cofactor analogue sinefungin (Wyszynski *et al.*, 1993; Klimasauskas and Roberts, 1995; Yang *et al.*, 1995). That AdoMet is bound after DNA can be easily accommodated by the structure since the cofactor-binding pocket is solvent-accessible and the pathway appears to be sufficiently large to allow AdoMet to enter the binding pocket even in the presence of DNA (Klimasauskas *et al.*, 1994).

Binding of the methyltransferase to substrate DNA results in large conformational changes of both the DNA and the protein. The DNA is bound in the cleft between the two domains, with the major groove facing the recognition domain and the minor groove facing the catalytic domain. Sequence-specific DNA binding is achieved via two glycine-rich (GKGGQGER and TLSAYGGGI) surface loops in the recognition domain that contact the DNA from the major groove side. Since these sequence specific contacts should be maintained during the reaction, the target cytosine can flip out only through the minor groove, and therefore towards the catalytic domain. While the overall conformation of the DNA remains undisturbed, the flipping of both the base and sugar of the target cytosine from the helix is accompanied by a substantial distortion (4 Å) of the phosphodiester backbone on the same strand, increasing the intrastrand phosphorus-phosphorus distances and providing a pathway for looping the target cytosine out of the helix. Extensive contacts between the methyltransferase and six phosphates surrounding the target cytosine (three on each strand) could contribute to this localised strand-opening (see fig. 1g).

In conjunction with the expulsion of the target cytosine from the DNA helix, Gln237 of the first glycine-rich loop of the recognition domain penetrates into the DNA duplex from the major-groove side, forming hydrogen bonds with all the polar groups of the unpaired guanine. In a similar move a large flexible loop from the catalytic domain (catalytic loop) containing the invariant Pro-Cys dipeptide motif common to all C5-methyltransferases moves nearly 25 Å to approach the DNA from



the minor-groove side. Ser87 from the loop hydrogen-bonds with and stabilises the position of Gln237, thereby locking the target cytosine in its extrahelical position, while at the same time the sulfhydryl group of Cys81 is brought into close proximity to the target cytosine. As a consequence of these conformational changes, the highly conserved amino acids Pro80 and Cys81 (motif IV), Glu119 (motif VI), and Arg165 (motif VIII) rearrange to form the active site.



**Fig. 1g: Schematic representation showing the specific base and phosphate contacts between *M.HhaI* and DNA.**

The recognition bases and contacted phosphates are shaded. Base contacts are shown with a thick line, phosphate contacts with a thin line, and contacts through main chain atoms with a dashed line. The encircled "W" indicates the water mediated contacts. (adapted from Klimasauskas *et al.*, 1994)

Refined crystal structures of *M.HhaI* with a variety of substrates in the presence or absence of cofactor offer a detailed picture of the active site at individual stages of the catalytic cycle, and both support and extend the proposed mechanism for C5-cytosine methylation (see fig. 1e). Once the target cytosine flips, it is held in place by a number of specific interactions with residues highly conserved among C5-methyltransferases (Phe79, Cys81, Ser85 of motif IV; Glu119 of motif VI; Arg165 of motif VIII). The thiol group of Cys81 is in close proximity to the C6 position of the flipped cytosine, while the methyl group of AdoMet lies close to the C5 position on the opposite side of the cytosine ring. The reaction is initiated by



the nucleophilic attack of Cys81 onto cytosine C6, while the observed hydrogen bond between Glu119 and the N3 of the cytosine ring provides a mechanism for the proposed N3 protonation which either occurs beforehand to positively polarise cytosine C6 and predisposes the ring to nucleophilic attack (Baker *et al.*, 1988) or concurrently to avoid the formation of a high energy carbanion intermediate (Chen *et al.*, 1993; Erlanson *et al.*, 1993). The resulting enamine has an increased reactivity at the C5 position, and attack on the methyl group of AdoMet by C5 leads to the transfer of the methyl group to the C5 position of the cytosine ring. Abstraction of the proton at C5 by  $\beta$ -elimination either via an activated water molecule or the neighbouring thioether S atom of Cys81 (Verdine, 1994) again yields an enamine, which undergoes  $\beta$ -elimination to yield the final product 5-methyl-cytosine.

How the reaction proceeds and the methylated cytosine is flipped back is unknown, but the methyl group present at the C5 position of the flipped cytosine after methylation substantially decreases the stability of the ternary complex of *M.HhaI* with methylated DNA and AdoHcy, and complexes containing fully methylated DNA dissociate rapidly in solution (Klimasauskas and Roberts, 1995; Yang *et al.*, 1995). This may at least in part be due to the irregular geometry adopted by the methyl group on C5, as is evident from crystals of *M.HhaI* with fully methylated DNA and AdoHcy. While the C5 of 5-methyl-cytosine in the complementary DNA strand adopts regular  $sp^2$  geometry and its methyl group is located in the plane of the cytosine ring, the methyl group of the flipped cytosine is located out of the plane of the cytosine ring and is directed towards the sulphur atom of AdoHcy. Since the position of the  $C_\alpha$  of Pro80 prohibits the methyl group from adopting a planar conformation, the resulting steric tension might explain the decreased stability of the complex.

It is important to note that crystal structures obtained in the presence of either unmethylated or methylated DNA (O'Gara *et al.*, 1996) show that both cytosine and 5-methyl-cytosine flip out of the DNA helix and fit into the active site of the enzyme. The methyltransferase therefore does not distinguish between substrate and product at the level of flipping the nucleotide out of the helix. Furthermore the crystal structure of *M.HhaI* complexed with methylated DNA in the presence of AdoHcy cannot accommodate an additional methyl group between the methylated cytosine and the sulphur atom of AdoHcy. This strongly points to the importance of AdoMet in the distinction between unmethylated (or hemi-methylated) and methylated DNA by methyltransferases, something which cannot be mimicked by AdoHcy.



The presence of most of the motifs conserved in all C5 methyltransferases in the catalytic domain of *M.HhaI* and their involvement in the binding of the cofactor AdoMet, binding of the flipped cytosine, and the catalytic process make it probable that all C5 methyltransferases share a common catalytic domain structure and catalytic process. Alignments of variable regions from methyltransferases recognising similar target sequences (see figure 1h), *e.g.* 5'-CCNGG-3' (*M.DsaV*, *M.ScrFI*, *M.SsoII*) and 5'-CCWGG-3' (*M.NlaX*, *M.EcoRII*), reveal segments with significant similarity, suggesting a similar structural framework for the protein-DNA interactions in these enzymes (Gopal *et al.*, 1994). However that little conservation is found between the variable regions of C5 methyltransferases recognising different targets might indicate different modes of target recognition.

This is illustrated by the recently elucidated crystal structure of *M.HaeIII* in complex with DNA (Reinisch *et al.*, 1995) which shows strong similarities to *M.HhaI*. Both flip out their target cytosine into active site pockets located in the large catalytic domains, which have a similar structural organisation and topology. The same arrangement of catalytic residues occurs in the two structures, and the same contacts to the substrate cytosine are made, in good agreement with the high degree of sequence conservation (34% identity) seen between the two catalytic domains. Despite the lack of any obvious sequence similarity between *M.HhaI* and *M.HaeIII* in the small domain, structural conservation is also found in two small regions in the recognition scaffold that supports the residues that interact with DNA. In contrast to the *M.HhaI*-DNA complex however, in which the DNA essentially maintains its B-form conformation, the DNA in the *M.HaeIII*-DNA complex is severely distorted, and local rearrangements in the pairing of the bases lead to the unstacking of bases and the opening of an 8 Å wide gap in the DNA. Again, alignments of *M.HaeIII* with other methyltransferases that recognise 5'-GGCC-3' reveal significant sequence similarity in the variable region between these enzymes (see fig. 1h).

The similarities between different C5 methyltransferases with respect to sequence conservation and catalytic mechanism is mirrored in the published structures. However no explanation is forthcoming for the different orders of binding of the substrates (*e.g.* *M.BsuRI*, Gunthert *et al.*, 1981) and different effects of the presence of AdoMet on DNA binding (*e.g.* *M.MspI*, Dubey and Roberts, 1992) reported for other C5 methyltransferases, or for the fact that while some obey Michaelis-Menten kinetics with respect to both AdoMet and DNA (*e.g.* *M.AluI*, Yoon *et al.*, 1985), others do not (*e.g.* *M.Bst153I*, Levy and Welker, 1981).



**Fig. 1h: Alignments of variable regions from methyltransferases recognising similar target sequences**

**a) Alignment of methyltransferases recognising 5'-CCNGG-3' or 5'-CCWGG-3'**

M.StyD4I	CCNGG	V	G	D	I	L	E	K	S	V	D	N	K	Y	T	L	S	D	A	L	W	N	G	H	Q	R	R	K	L	V	N	A	A	A	G	K	G	F	G	Y	G	L	F	N	E	N		S	P	Y	T	N	T	I	S	A	R	Y	Y	K	D	G	S	E	I	L	I	E	Q	
M.ScrFI	CCNGG	V	G	N	I	L	E	S	V	V	D	D	K	Y	T	I	S	D	K	L	W	D	G	H	Q	R	R	K	T	E	N	K	K	N	G	K	G	F	G	Y	T	L	F	N	Q	D		S	E	Y	T	N	T	I	S	A	R	Y	Y	K	D	G	S	E	I	L	I	E	Q	
M.DsaV	CCNGG	V	G	D	L	L	E	K	E	V	D	E	K	Y	T	I	T	D	R	M	W	E	G	H	Q	N	R	K	K	A	H	R	K	R	G	N	G	F	G	F	S	L	V	N	R	N		S	S	Y	T	R	T	I	S	A	R	Y	Y	K	D	G	S	E	V	L	V	E	Q	
M.SsoII	CCNGG	V	G	D	I	L	E	K	S	V	D	N	K	Y	T	L	S	D	A	L	W	N	G	H	Q	R	R	K	L	V	N	A	A	A	G	K	G	F	G	Y	G	L	F	N	E	N		S	P	Y	T	N	T	I	S	A	R	Y	Y	K	D	G	S	E	I	L	I	E	Q	
M.NlaX	CCWGG	V	G	D	I	L	E	A	Y	P	D	E	K	Y	T	I	S	D	K	L	W	Q	G	Y	Q	R	R	K	A	E	N	R	A	A	G	K	G	F	G	Y	G	L	F	N	A	E		S	A	Y	T	N	T	I	S	A	R	Y	Y	K	D	G	S	E	I	L	I	E	Q	
M.EcoRII	CCWGG	F	G	E	L	L	E	P	V	V	D	S	K	Y	I	L	T	P	K	L	W	E	Y	L	Y	N	Y	A	K	K	H	A	A	K	G	N	G	F	G	F	G	L	V	N	P	E	N	K	E	S	I	A	R	T	L	S	A	R	Y	H	K	D	G	S	E	I	L	I	D	R
dcm	CCWGG	L	A	Q	L	L	D	P	M	V	E	A	K	Y	I	L	T	P	V	L	W	K	Y	L	Y	R	Y	A	K	K	H	Q	A	R	G	N	G	F	G	Y	G	M	V	Y	P	N	N	P	Q	S	V	T	R	T	L	S	A	R	Y	Y	K	D	G	A	E	I	L	I	D	R

**b) Alignment of M.Hae III with other methyltransferases recognicing 5'-GGCC-3'**

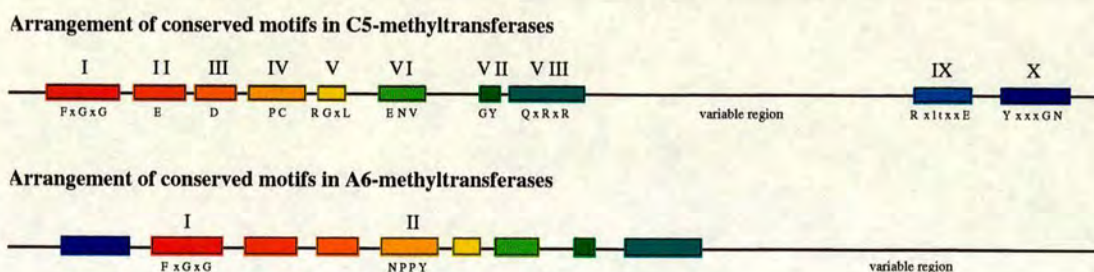
DNA contacts in HaeIII			+		*	+	*		*	*	*		+		+		+	*	*	*																													
M.Hae III	211-258	H	E	Y	F	I	G	S	Y	S	T	I	F	M	S	R	N	R	V	R	Q	W	N	E	P	A	F	T	V	Q	A	S	G	R	Q	C	Q	L	H	P	Q	A	P	V	M	L	K	V	S
M.Fnu DI	214-261	H	E	Y	L	T	G	S	Y	S	T	I	F	M	S	R	N	R	V	R	Q	W	E	Q	P	A	F	T	V	Q	A	S	G	R	Q	C	Q	L	H	P	Q	A	P	T	M	I	K	I	D
M.Mth TI	212-259	H	E	Y	M	T	G	T	F	S	S	R	Y	M	S	R	N	R	V	R	S	W	D	E	V	S	F	T	I	Q	A	G	G	R	H	A	P	C	H	P	Q	A	N	K	M	I	K	V	G
M.Ngo PII	226-273	N	E	Y	F	T	G	S	F	S	P	I	F	M	S	R	N	R	V	K	A	W	D	E	Q	G	F	T	V	Q	A	S	G	R	Q	C	Q	L	H	P	Q	A	P	K	M	E	K	H	G
M.Bsu RI	235-332	G	P	Y	F	T	G	S	Y	S	S	I	Y	M	S	R	N	R	K	K	S	W	D	E	Q	S	F	T	I	Q	A	S	G	R	Q	A	P	L	H	P	G	G	L	S	M	K	K	I	G
M.Bsp RI	234-281	G	P	Y	F	T	G	S	Y	S	T	I	F	M	S	R	N	R	K	K	K	W	T	D	Q	S	F	T	I	Q	A	S	G	R	Q	A	P	I	H	L	G	G	L	P	M	E	K	V	D

**red** : invariant residues  
**blue** : 80% conserved or only conservative substitution  
**boxed** : scaffold region  
**grey** : important for DNA recognition in M.HaeIII  
+ : contact with DNA backbone  
\* : involved in base specific recognition



## A6-Methyltransferases

Although early sequence alignments identified only two conserved motifs in A6- and C4-methyltransferases (Chandrasegaran and Smith, 1988; Klimasauskas *et al.*, 1989), more exhaustive sequence comparisons (this thesis; Noyer-Weidner *et al.*, 1994; Dryden *et al.*, 1995) revealed a total of nine conserved motifs, corresponding to the motifs I to VIII and X previously defined in C5-cytosine methyltransferases. The biggest difference between C5-methyltransferases and A6-methyltransferases of the  $\gamma$ -class is in regard to the location of the variable region in the primary sequence (between motifs VIII and IX in the C5-methyltransferases, C-terminal in A6-methyltransferases of the  $\gamma$ -class, see fig. 1i), requiring either one or two connections between the catalytic and the recognition domains.



**Fig. 1i: Location of variable region and conserved motifs in the primary sequence of C5- and A6-methyltransferases**

The structural conservation of these motifs was confirmed by the elucidation of the *M.TaqI* crystal structure in complex with its cofactor AdoMet (Labahn *et al.*, 1994, see fig. 1j). *M.TaqI* is a bilobal protein in which two domains of about equal size are arranged in the form of a C with a wide cleft suitable to accommodate the DNA substrate. Comparisons of this structure with the previously determined *M.HhaI* (Cheng *et al.*, 1993) and catechol O-methyltransferase (Vidgren *et al.*, 1994) structures show a remarkably similar catalytic domain with an alpha/beta structure containing a mixed central beta-sheet, in which the cofactor binding sites are almost identical and the essential catalytic amino acids coincide (Malone *et al.*, 1995; O'Gara *et al.*, 1995; Schluckebier *et al.*, 1995a).





**Fig. 1j: *M.TaqI* structure in complex with AdoMet**

The two domain structure of the *TaqI* methyltransferase is readily recognisable, and DNA is expected to be bound in the cleft between the two domains. The methylation domain has a structure very similar to the methylation domain of the C5 methyltransferases *M.HhaI* and *M.HaeIII*.

Molecule	Part	Display	Colour
<b><i>M.TaqI</i></b>	$\alpha$ -Helices	ribbon	magenta
	$\beta$ -strands	ribbon	yellow
	turns	ribbon	pale blue
	other residues	ribbon	white
<b>cofactor</b>	AdoMet	spacefill	CPK

Although no structure of *M.TaqI* in complex with DNA is available, several lines of evidence point to the unstacking and flipping-out of the target adenine. While undistorted B-DNA could be modelled into the available structure, the distance of 15 Å between the methyl donor AdoMet and the target adenine is too far for a direct transfer, and there are no obvious side chains in the vicinity that would be capable of assisting in the reaction. A gap in the catalytic domain is suited to accommodate a flipped-out adenine, and the dependence of the rates of methylation on the nature of the 3'-neighbouring base agrees with the unstacking of adenine and its 3'-neighbour,



but not a distortion of both strands (Woelcke and Weinhold, 1995). However, the distance of the AdoMet methyl group to adenine N6 after base-flipping would still be 8 Å, and a torsion of the cofactor similar to the one seen in *M.TaqI* in complex with the cofactor analogue sinefungin appears to be required to bring cofactor and substrate sufficiently close together. Alternatively or simultaneously a conformational change in the protein upon DNA binding may occur since three of the flexible loops that might be involved in DNA binding orientate the domains relative to each other. A precedent for this is seen in the interactions of *M.EcoRI* with its substrate, where spectroscopic data implicate conformational changes upon DNA binding (Maegley *et al.*, 1992), consistent with previous results from limited proteolysis studies (Reich *et al.*, 1991).

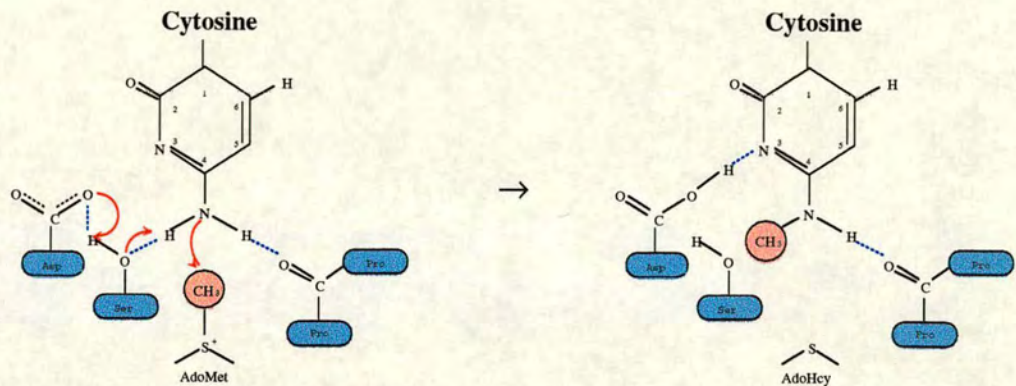
Little is known about the enzymatic mechanism of A6-methyltransferases. Contrary to the findings with C5-methyltransferases, no covalently bound intermediate could be trapped, and a well established methylation reaction in organic synthesis (Dimroth rearrangement) in which the methyl group is initially transferred to the N1 position of adenosine followed by isomerisation of the resulting N1-methylamino group with the 6-amino group to give the end product 6-methyladenine could be excluded (Pogolotti *et al.*, 1988). Since transfer of the methyl group in both enzyme-catalysed C5- and A6-methylation reactions involves inversion of configuration of the methyl group, and therefore necessitates an uneven number of transfer reactions, an intermediate transfer of the methyl group onto the enzyme seems unlikely (Ho *et al.*, 1991) and a single, direct transfer of the methyl group from AdoMet onto the target adenosine is suggested (see fig. 1k). Possible hydrogen bonds of the terminal carbonyl group of Asn105 of the NPPY motif in *M.TaqI* to the N6 (NH<sub>2</sub> group) of a flipped adenine and interactions of the main-chain oxygens of the two proline residues with the amine hydrogens would negatively polarise N6 and enable a direct attack by the CH<sub>3</sub><sup>+</sup> of AdoMet.

Like C5-methyltransferases, A6-methyltransferases display a variety of effects in their interactions with substrate and cofactor. Spectroscopic data for *M.EcoRI* imply conformational changes upon DNA binding but not upon AdoMet binding (Maegley *et al.*, 1992), and kinetic data support a mechanism in which the methyltransferase binds AdoMet and noncanonical DNA randomly but where recognition of the canonical site requires AdoMet to be bound (Reich and Mashhoon, 1991). Data for *dam* methylase require the binding of AdoMet to two nonidentical sites, one allosteric, the other catalytic (Bergerat and Guschlbauer, 1990; Bergerat *et al.*, 1991), and are consistent with the activation of DNA binding by AdoMet and a fit-induced rearrangement of

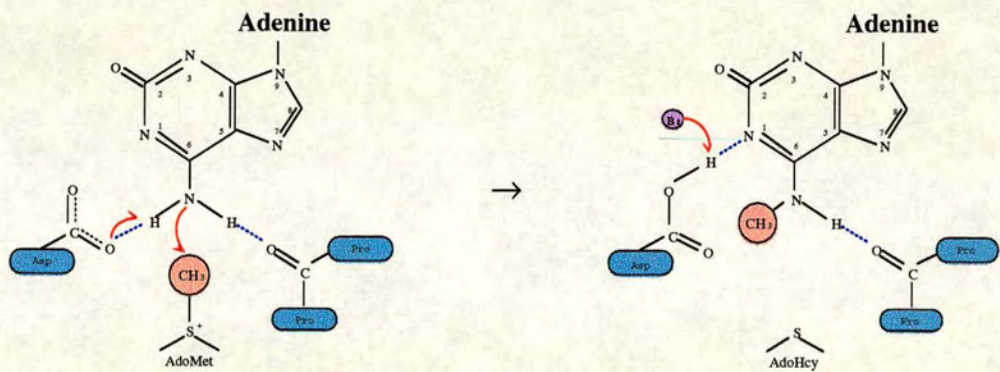


**Fig. 1k: Proposed reaction mechanism for amino-methyltransferases**

**with SPPY/F in motif IV:**



**with DPPY/F in motif IV:**



**with NPPY/F in motif IV:**

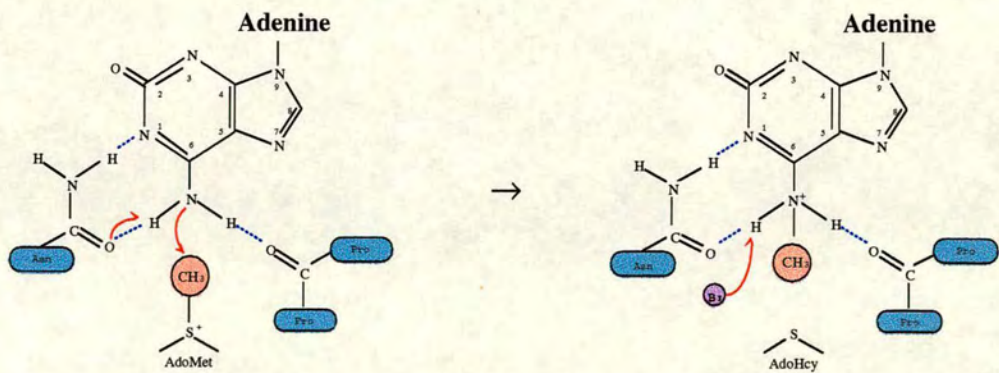


Figure adapted from Gong *et al.*, 1997. A general base (B:), which could be a water molecule, might be needed to eliminate the proton.



the initial ternary complex upon binding to the cognate site. *M.EcaI* catalyses the transfer of a methyl group from AdoMet to adenine with a random rapid-equilibrium mechanism, and recognition of the specific site of DNA occurs after the binding of AdoMet (Szilak *et al.*, 1993).

### C4-Methyltransferases

Very little is known about C4-methyltransferases, the smallest group among the known DNA methylases. Since both A6- and C4-methyltransferases methylate an exocyclic amino group, similarities in the catalytic mechanism are expected, but the high degree of sequence similarity to the A6-, but not to C5-methyltransferases came as somewhat of a surprise. The most obvious difference between C4- and A6-methyltransferase sequences is a conserved Serine instead of Asparagine or Aspartate in the NDPPYF motif, although this does not represent an essential functional difference as the C4-methyltransferase *M.BamHI* has DPPF instead.

The recently solved *M.PvuII* crystal structure (Gong *et al.*, 1997) shows strong similarities to those of *M.HhaI*, *M.HaeIII*, and *M.TaqI* with regard to the AdoMet-binding domain. This is consistent with a model predicting that all Type II DNA methyltransferases share a common structural fold while having the major functional regions permuted into three distinct linear orders, corresponding to the three subgroups  $\alpha$ ,  $\beta$ ,  $\gamma$  (Malone *et al.*, 1995). The AdoMet-protein interactions are almost identical, and structural similarities among the active sites reveal that catalytic amino acids essential for cytosine N4 and adenine N6 methylation coincide spatially with those for cytosine C5 methylation.

If flipped out of the double helical DNA the target base would fit into the concave active site next to AdoMet. By assuming a base flipping mechanism, modelling of an appropriate DNA molecule into the methyltransferase structure, and superimposition of the common  $\alpha$ -helix/ $\beta$ -sheet structures, it has been possible to infer the interactions of the flipped base with both C4- and A6-methyltransferases (see fig. 1k). The model suggests that for C4 methyltransferases methylation of the exocyclic amino group results from a direct attack of the activated cytosine N4 on the AdoMet methyl group, in analogy with the previously proposed mechanism for DNA adenine methylation (Pogolotti *et al.*, 1988; Ho *et al.*, 1991; Schluckebier *et al.*, 1995a).



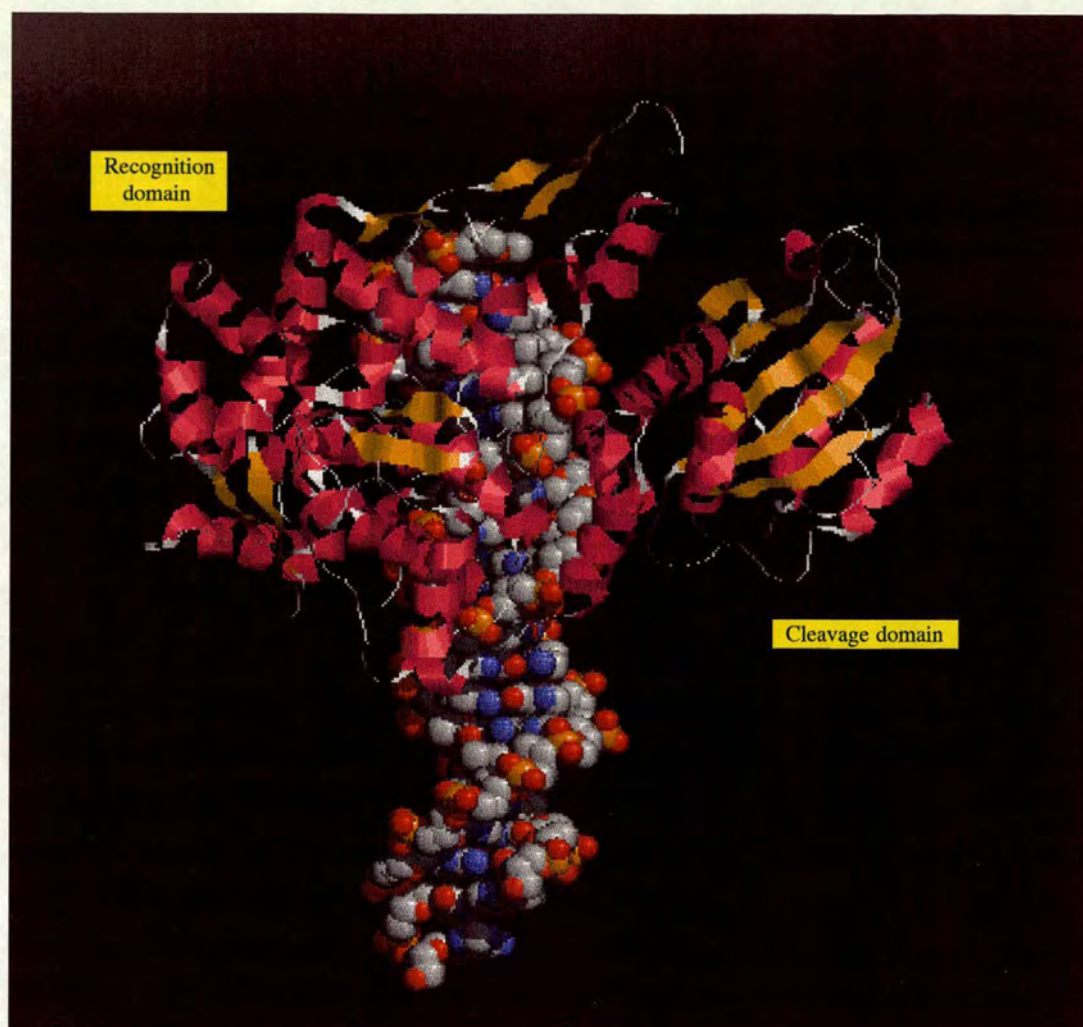
### 1.4.2. Type IIs enzymes

The relative simplicity of Type II endonucleases was a major contributing factor for their successful application to the manipulation of DNA. However the tight integration of DNA binding and restriction functions (*i.e.* the lack of a distinct domain structure) combined with our limited understanding of how sequence specific recognition is achieved in the first place renders them unsuitable for manipulation (except for the relaxation of specificity). Consequently their use is dependent on the availability of enzymes with suitable specificities or the construction of appropriate target sequences in the DNA of interest.

Although Type IIs endonucleases share several characteristics with Type II endonucleases - chiefly among them similar cofactor requirements and a simple relationship between the recognition and cleavage site - they achieve a whole new level of complexity by separating target recognition and cleavage functions into distinct domains as exemplified by the best studied Type IIs system *FokI* (Li *et al.*, 1992; Skowron *et al.*, 1993; Waugh and Sauer, 1993; Li, 1994; Hirsch *et al.*, 1997; Wah *et al.*, 1997).

These domains interact with two discrete sites on double-stranded DNA - the sequence-specific recognition site and an unspecific cleavage site - and seem to function independently from each other (Li *et al.*, 1992). As a consequence the target recognition domains of Type IIs restriction endonucleases can be exchanged against other DNA binding domains (Kim and Chandrasegaran, 1994; Huang *et al.*, 1996; Kim *et al.*, 1996; Schaeffer *et al.*, 1996), thereby constructing restriction enzymes with new specificities. Changes in the spacing between recognition and cleavage site can be achieved by insertions into and deletions from the proposed linker region between the domains (Kim *et al.*, 1994; Li, 1994). The discovery of single amino acid substitutions which uncouple the DNA binding and strand scission activities of the *FokI* endonuclease and completely eliminate its ability to cleave either strand of substrate DNA indicate the use of a single catalytic centre for the cleavage of both strands of DNA (Waugh and Sauer, 1993). As anticipated the recent elucidation of the R.*FokI* crystal structure in complex with DNA (see fig. 11; Wah *et al.*, 1997) confirmed that the enzyme contains amino- and carboxy-terminal domains corresponding to the DNA-recognition and cleavage functions, respectively.





**Fig. 11: Crystal structure of *R.FokI* in complex with DNA**

The predicted two domain structure of the *FokI* endonuclease is readily recognisable in the crystal structure. While the recognition domain contacts the DNA, the cleavage domain is sequestered alongside the recognition domain.

Molecule	Part	Display	Colour
<b><i>R.FokI</i></b>	$\alpha$ -Helices	ribbon	magenta
	$\beta$ -strands	ribbon	yellow
	turns	ribbon	pale blue
	other residues	ribbon	white
<b>DNA</b>	double helix	wireframe	CPK



The structure shows that the recognition domain is composed of three smaller subdomains (D1, D2 and D3) which are related to the helix-turn-helix domain of the CAP protein (catabolite gene activator protein). While subdomain D1 recognises base pairs at the 3' end of the recognition sequence (GGATG), subdomain D2 recognises base pairs at the 5' end of the recognition sequence (GGATG). Surprisingly subdomain D3 - which is most similar to CAP and related proteins - barely touches the DNA and instead has been co-opted for protein-protein interactions with the cleavage domain, with its recognition helix being used primarily to piggyback the cleavage domain onto the recognition domain.

The cleavage domain has a structure remarkably similar to a subunit of the dimeric endonuclease *Bam*HI, the two sharing a similar  $\beta$ -sheet core surrounded by  $\alpha$ -helices on both sides and superimposing with a r.m.s.d. of 2.0 Å (using 61 C $\alpha$  atoms). Consequently the cleavage domain contains only a single catalytic centre, and in order for the enzyme to cleave both strands it would have to either dimerise on the DNA or the cleavage domain would have to flip between two orientations, cleaving first one DNA strand and then the other.

The positioning of the cleavage domain alongside the recognition domain prevents it from associating with the DNA cleavage site. This agrees well with footprinting studies on *R.FokI* with DNA which show a lack of protection at the cleavage site (Li *et al.*, 1993; Waugh and Sauer, 1993; Yonezawa and Sugiura, 1994). The sequestering of the cleavage domain may also explain how *R.FokI* manages to regulate its cleavage activity until it is required. Upon finding a suitable restriction site a simple rotation around the linker segment would bring the cleavage domain to the major groove for cleavage.

Interference of a methyl group in the target site with the binding of *R.FokI* protects methylated or hemi-methylated DNA against cleavage (Skowron *et al.*, 1993). However an unusual class of mutants that cleave hemi-methylated DNA have been found (Waugh and Sauer, 1994), and Waugh and Sauer postulate that there may be a rate-limiting step in the *R.FokI* cleavage reaction that involves the dissociation of the cleavage domain from the recognition domain. Consequently mutations which occur at the interdomain interface where they would be expected to relax the domain-domain association would accelerate the transition and allow cleavage at hemi-methylated DNA sites. Consistent with this view seven of the eight mutations found so far are located at the interdomain interface in the crystal structure.



Modification is usually accomplished by a pair of methyltransferases, which unlike the endonucleases seem to be quite similar to their Type II counterparts. In the case of *FokI* the two adenine methyltransferases are fused together into a single, bifunctional enzyme (Landry *et al.*, 1989; Looney *et al.*, 1989), and the same seems to be true for some other systems like *Alw26I*, *Eco31I*, and *Esp3I*, where a single enzyme protein displays both C5- and A6-methyltransferase activity (Bitinaite *et al.*, 1992). Consistent with this sets of conserved motifs characteristic for both A6- and C5-methyltransferases occur in the primary sequence of M.*Alw26I*.

Two other Type II systems - *MboII* and *NgoBI* - seem to modify only one strand of the recognition site, reminiscent of Type III enzymes (Piekarowicz *et al.*, 1988; Bocklage *et al.*, 1991), and it is as yet unclear how they protect chromosomal DNA during replication against cleavage by the restriction endonuclease. However the same was suspected for the *HphI* system (Szybalski *et al.*, 1991) before the identification of a second methyltransferase (Lubys *et al.*, 1996).

### 1.4.3. Type I enzymes

Type I enzymes show the highest degree of modularity among the known R/M-systems, for different functions are located on and performed by different subunits. But contrary to Type II systems where the endonuclease and methyltransferase act independently of each other and AdoMet and DNA are mere substrates in the respective reactions, Type I enzymes closely integrate the different functions with each other so that restriction proficiency cannot be achieved without the assembly of a methylation-competent complex first. Which reaction finally takes place depends on the nature of the substrate, and both AdoMet and DNA act as allosteric effectors in the determination of the reaction pathway (Bühler and Yuan, 1978; Burckhardt *et al.*, 1981a; Burckhardt *et al.*, 1981b).

#### 1.4.3.1. The subunits

##### HsdS

DNA binding is mediated by the HsdS subunits of the Type I R/M systems which, reminiscent of the target recognition domains of Type II enzymes, confer sequence specificity to their respective enzymes. However the bipartite and interrupted



nature of their target sequence is reflected in the modular structure of the *hsdS* genes, where two large variable regions alternate with smaller regions which are highly conserved within a family (see fig. 1m). The discovery (Bullas *et al.*, 1976) and characterisation (Fuller-Pace *et al.*, 1984; Nagaraja *et al.*, 1985; Fuller-Pace and Murray, 1986) of a system with a new recognition specificity that arose through recombination between the *hsdS* genes of two Type IA systems, and the subsequent construction of more hybrids in both Type IA and IC systems (Gann *et al.*, 1987; Gubler *et al.*, 1992) clearly demonstrated that the 5' variable region dictates recognition of the trinucleotide component of the target sequence, while the 3' variable region governs recognition of the tetra- or penta-nucleotide component.

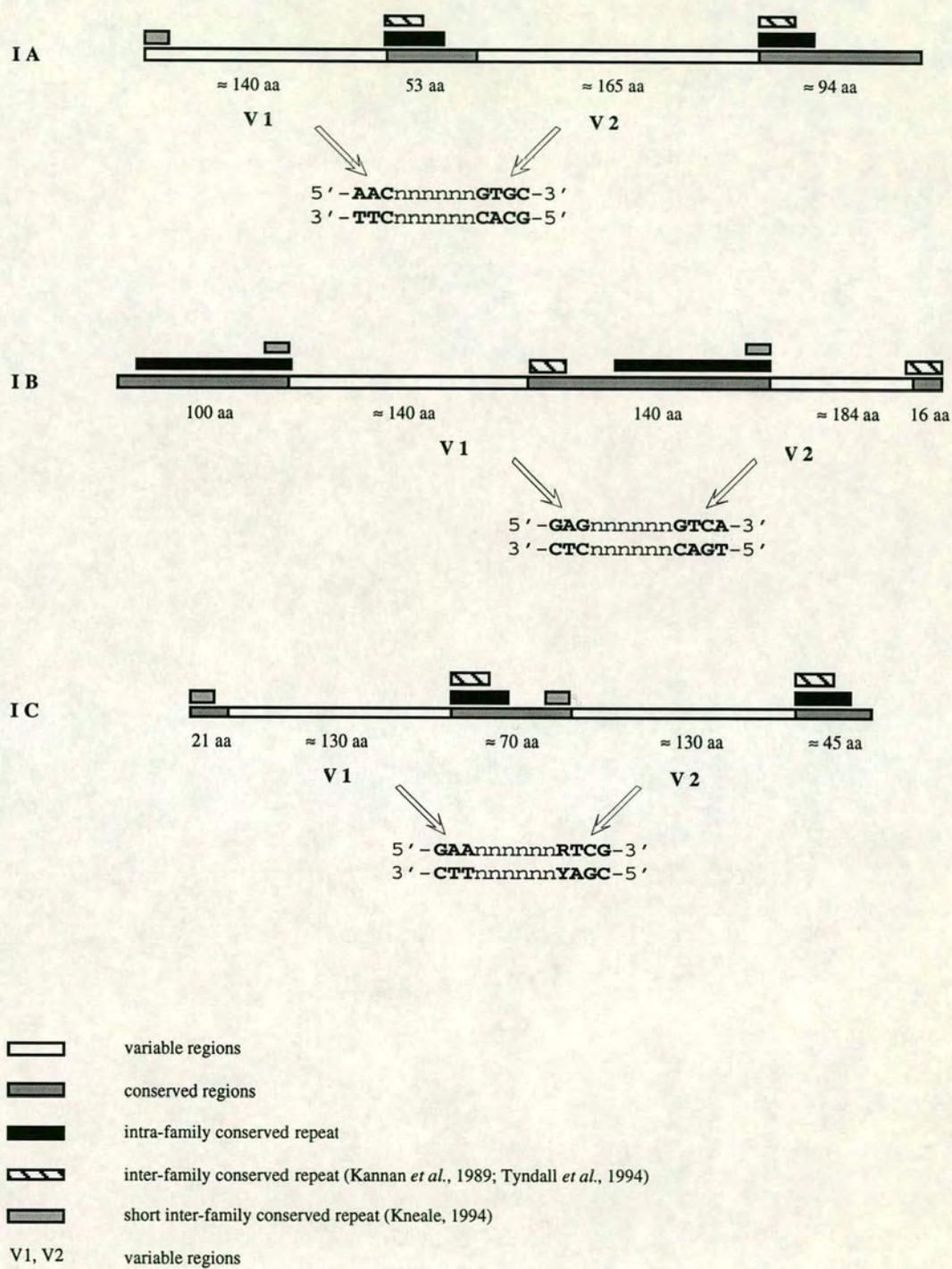
That the variable regions encode two independent target recognition domains (TRD) is consistent with the results of sequence comparisons between the deduced amino acid sequences of different HsdS subunits (Fuller-Pace and Murray, 1986; Cowan, 1989; Cowan *et al.*, 1989; Kannan *et al.*, 1989; Skrzypek and Piekarczyk, 1989). While systems that recognise dissimilar targets share little similarity in the variable regions, sequences that recognise the same trinucleotide component show strong similarities in the N-terminal variable region, even if the systems belong to different families.

The conserved regions are thought to interact with two HsdM subunits to form the trimeric methyltransferase. Sequence analysis of the deduced amino acid sequences revealed the presence of sequence repeats between the conserved regions (intrafamily repeats) and showed a circular permutation in the arrangement of variable and conserved regions between families (see fig. 1m; Argos, 1985; Kannan *et al.*, 1989; Kneale, 1994), in which part of the conserved region is located at the C-terminus and continued at the N-terminus. On the basis of this "split repeat", a circular organisation of the domains in the HsdS subunit was proposed (see fig. 1n) in which one M-subunit interacts with the N- and C-terminal conserved regions of the S-subunit, which are thereby brought into close proximity, while the second M-subunit makes equivalent contacts with repeated sequences in the central conserved domain.

Experimental support for this model comes from deletion mutants of *hsdS*, which indicate that an amino-TRD can function as a carboxy-TRD and *vice versa* (Piekarczyk *et al.*, 1985; Abadijewa *et al.*, 1993; Meister *et al.*, 1993; Price *et al.*, 1993; MacWilliams and Bickle, 1996). Deletion mutants that lack the 3' half of the *hsdS* gene - *i.e.* express only the first TRD of HsdS - code for a methyltransferase in which the recognition sequence is an inverted repeat of the trimeric component of



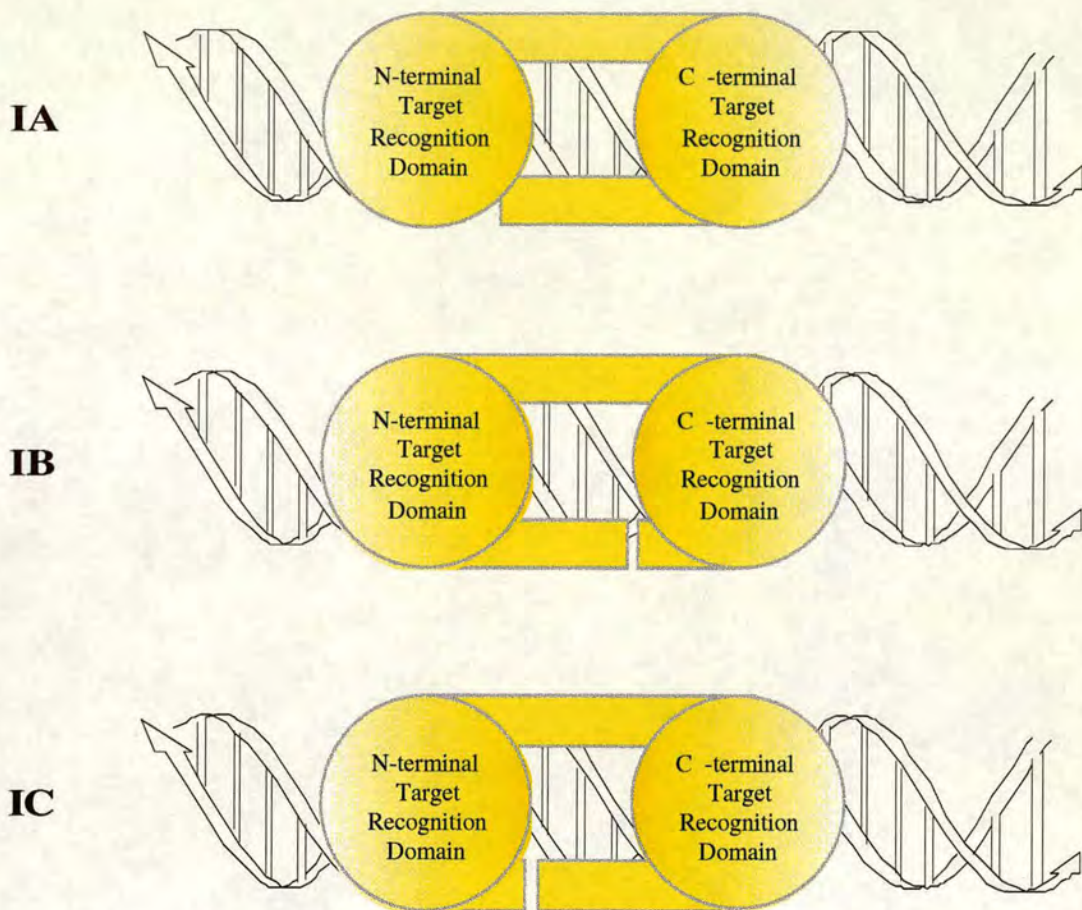
# Family



**Fig. 1m: Modular structure of the *hsdS* genes and proteins**

The figure shows the location and approximate size of various conserved and variable regions within the *hsdS* genes of the IA, IB, and IC family (represented by *hsdS-K*, *hsdS-E* and *hsdS-R124* respectively) as well as the number of amino acids they encode. V1 encodes the TRD recognising the trinucleotide compartment of the recognition site, while V2 encodes the TRD recognising the tetranucleotide part.





**Fig. 1n: Circular organisation of the *HsdS* domains**

Although the relative size and location of conserved and variable regions in the *hsdS* genes varies between families, the proteins they encode are expected to share the same general structure. Of special importance is that the asymmetric target sequence is recognised in a pseudo-symmetrical manner, thereby enabling the assembly of a symmetrical protein complex on an asymmetrical target. It also suggests that the two target recognition domains read their respective half-targets in opposite directions.

the original recognition sequence. Correspondingly deletions of the 5' half of the gene (*i.e.* only encoding the second TRD of HsdS), when fused to the appropriate transcriptional and translational start signals, code for a system where the recognition sequence is an inverted repeat of the tetrameric component of the original recognition sequence. Co-expression of the 5' *hsdS* mutant and the 3' *hsdS* mutant along with *hsdM* regenerates the wild-type methylation specificity, suggesting a free assortment of the mutant "half-subunits" in the cell, where the mutant methyltransferase is assembled from two inverted HsdS half-subunits and possibly held together by the HsdM subunits. As a consequence the methyltransferase would interact with DNA in a pseudo-symmetric fashion, in which the trinucleotide part of the recognition sequence would be read in one direction, while the corresponding tetra- or pentanucleotide part



of the recognition sequence would be read in the opposite direction (Sturrock, 1997). The recent characterisation of the *StySKI* system (Thorpe *et al.*, 1997), which recognises the unusual target sequence 5'-CGAT(N<sub>7</sub>)GTTA-3', lends strong support to this view since its amino-TRD shows 35% amino acid identity to the carboxy-TRD of *EcoR124I* which recognises the complementary, but degenerate, DNA sequence 5'-RTCG-3'.

The *hsdS* genes of the recently discovered ID family, which includes enzymes from *Salmonella enterica* (Titheradge *et al.*, 1996), *Pasteurella haemolytica* (Highlander and Garza, 1996), and a predicted open reading frame from *Haemophilus influenzae* (Fleischmann *et al.*, 1995), share a strongly conserved region with each other but lack an internal repeat sequence, and it is as yet unclear whether they interact with their respective HsdM subunits in a novel way or whether they might interact with two different HsdM subunits to form a complex of the form M<sub>a</sub>SM<sub>b</sub>. The presence of two open reading frames with similarities to adenine methylases in the *Salmonella* system seems to point to the latter possibility.

The conserved regions of HsdS subunits are also involved in the spacing of the two target recognition domains, as is evident from studies on the *EcoR124I* and *EcoR124II* systems, both of which recognise a GAA trinucleotide and a RTCG tetranucleotide target. The only difference between the two systems is a 12 base-pair sequence that is repeated twice in the central conserved region of the *EcoR124* gene and three times in the *EcoR124II* gene (Price *et al.*, 1989). The four extra amino acids encoded by this additional repeat are sufficient to increase the spacing between the two target components from six base pairs in the *EcoR124I* system to seven base pairs in the *EcoR124II* system, which on B-form DNA would correspond to both an increase in distance by 0.34 nm and a 36 degree rotation of the two target sites relative to each other. How this is accommodated in the methyltransferase complex is unknown, especially since there is no corresponding change in the equivalent second conserved region, but the effects on restriction activity and modification specificity of single amino acid substitutions and small insertions and deletions in this repeat (Gubler and Bickle, 1991) strongly support the idea that the repeated amino acid motif forms part of a flexible interdomain linker.

Little is known about how sequence specific binding is achieved. The interchangeability of subunits within a family as demonstrated early on by complementation tests as well as the results of domain swapping experiments exclude a role for HsdM in the determination of sequence specificity. While it is therefore not



surprising that purified HsdS from the *EcoR124I* system binds specifically to its DNA target sequence (Kusiak *et al.*, 1992), the presence of the HsdM subunit stimulates DNA binding, possibly *via* specific interactions with its target adenine. It is noteworthy that HsdM subunit added subsequently does not appear to bind to the HsdS-DNA complex, indicating possible structural changes in the HsdS subunit upon binding to DNA. The fact that the adenosyl residues that are methylated in the modification reaction are always situated ten or eleven base-pairs apart suggests that the enzymes bind to DNA along one face of the double helix (Nagaraja *et al.*, 1985), and footprinting experiments with the *EcoKI* methyltransferase support the idea that the methyltransferase contacts the DNA in two successive major grooves (Powell and Murray, 1995). The availability of new sequence data of variable regions from a variety of sources as well as the realisation that the two TRD should be treated independently for use in sequence alignments and that the target sequence of the second TRD should be read in 3'→5' direction made more exhaustive sequence alignments and secondary structure predictions possible (Sturrock, 1997), and led to the prediction that all of the 51 known TRDs have the same tertiary structure and furthermore are similar to the TRD of the C5-methyltransferase *M.HhaI*.

## HsdM

The purification of HsdM from a strain containing only the *EcoKI hsdM* gene on an inducible plasmid (Winter, 1992) revealed the existence of two forms of the M subunit. It was thought at the time that the two peaks eluting from a gel filtration column represent monomeric and dimeric forms of HsdM, but they could also represent the free subunit and a binary complex between HsdM and AdoMet as seen in the purification of *HhaI* methyltransferase. It should be noted that these two explanations are not mutually exclusive. However the later explanation would also be consistent with results from a Blue Sepharose column, on which nucleotide binding proteins are retained, and where HsdM was found both in the eluate and in the flow-through. The biochemical characterisation of the purified HsdM from the *EcoR124I* system (Taylor, 1992) also supports the idea that this subunit is monomeric in solution.

The HsdM subunit confers AdoMet binding and methylation functions to the methyltransferase complex, but the close integration of its functions with the formation of a methylation competent complex does not allow a full range of functional assays with the HsdM subunit alone. UV-crosslinking experiments with radiolabelled



AdoMet and purified HsdM showed that AdoMet binding is an intrinsic property of the HsdM subunit itself (Powell *et al.*, 1993), and confirmed the existence of a cofactor binding site in HsdM which had been previously inferred from the presence of two motifs common to all adenine methyltransferases. Site-directed mutagenesis of the two motifs (Willcock *et al.*, 1994) substantiates their involvement in the enzymatic activities associated with the HsdM subunit. A mutation in the first motif (the presumed AdoMet binding site) abolishes cofactor binding and enzyme activity, and points to the existence of a single AdoMet binding site in HsdM. Mutations in the second motif reduce or abolish enzyme activity without affecting cofactor binding, and that a mutation in the second motif can enhance the UV-induced cross-linking between the enzyme and AdoMet suggests that both motifs are close in space to the methyl-group donor.

Analysis of the products of the limited digestion of the *EcoKI* methyltransferase by various proteases confines the interface on the HsdM subunit that mediates association with the HsdS subunit to a region between Thr280 and Glu433 of the primary sequence (Cooper and Dryden, 1994), and also defines a C-terminal tail region whose susceptibility to cleavage is dependent on the methylation state of the DNA.

## **HsdR**

The addition of the HsdR subunit to the methyltransferase complex not only adds restriction capabilities to the complex, but  $Mg^{2+}$  and ATP binding, ATP hydrolysis and DNA translocation activities as well. Consistent with this, amino acid sequence comparisons between subunits from all the families reveal two conserved regions in HsdR, one of which contains seven motifs typical of DNA and RNA helicases (Linder *et al.*, 1989; Gorbalenya and Koonin, 1991; Murray *et al.*, 1993; Fuller-Pace, 1994; Titheradge *et al.*, 1996) while the other conserved region contains a motif that is structurally conserved in the active site of Type II restriction endonucleases, implying that Type I enzymes might restrict DNA in a manner very similar to Type II enzymes.

The recently solved crystal structure of a DNA helicase from *Bacillus stearothermophilus* (Subramanya *et al.*, 1996) shows that all the motifs conserved among DNA and RNA helicases are located around the nucleotide-binding site, and indicates their involvement in ATP binding (motif I and parts of motifs IV, V, VI), DNA binding (motif Ia),  $Mg^{2+}$  binding (motif II, DEAD), coupling of ATP hydrolysis



to helicase activity (motif III), and helicase activity (motif V and VI). Based on their sequence similarity a similar arrangement can be expected for HsdR.

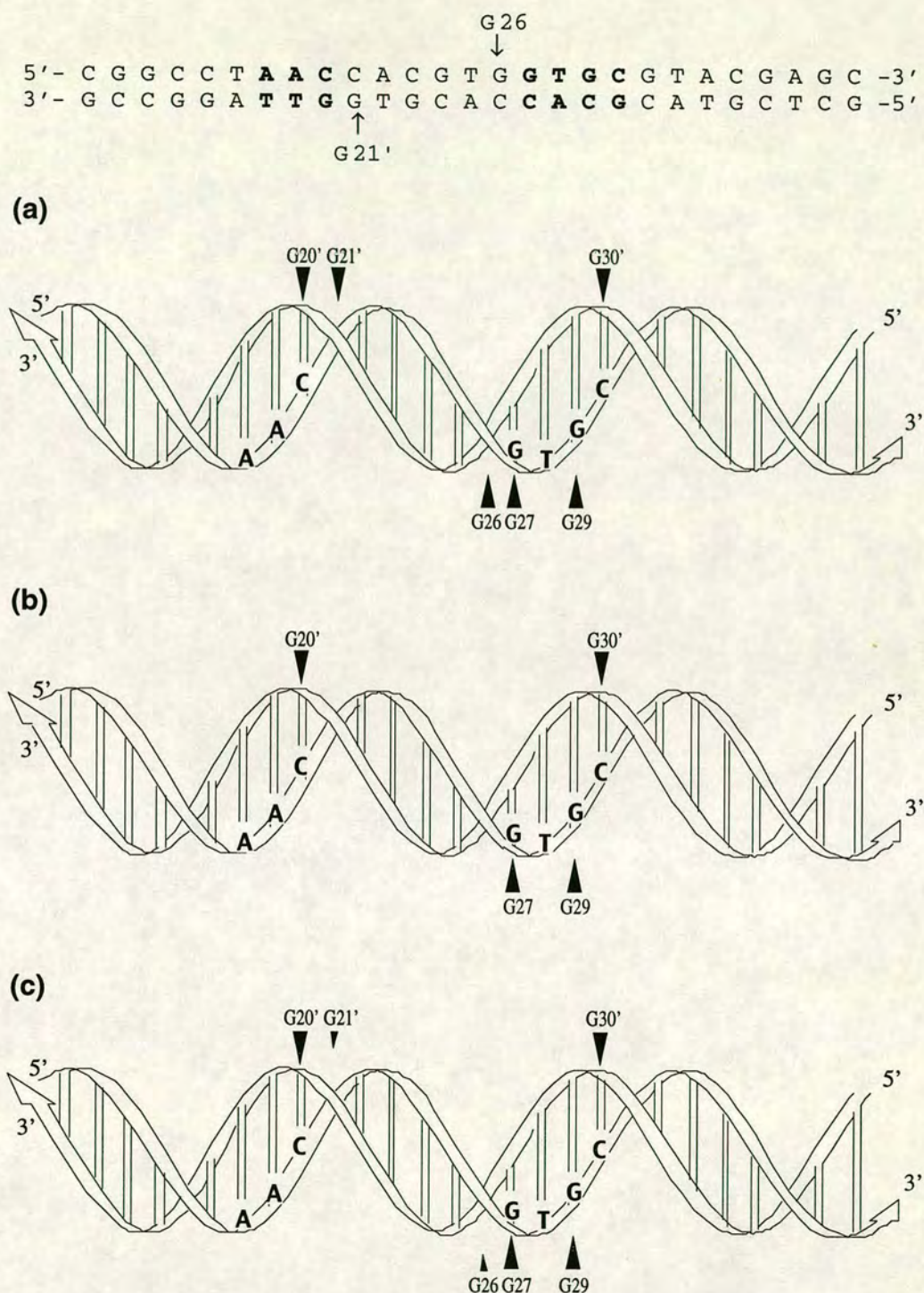
#### 1.4.3.2. The Methyltransferase

The purification of the *EcoKI* methyltransferase from an overexpressing strain yields not only the functional trimeric methyltransferase  $M_2S_1$ , but also free M and an inactive  $M_1S_1$  complex (Dryden *et al.*, 1993). The HsdM subunit combines with the HsdS subunit to form the functional methyltransferase, which in addition to the DNA binding properties of the HsdS subunit and the AdoMet binding functions contributed by the HsdM subunit displays additional capabilities which can be attributed to interactions of the subunits with each other and their substrates AdoMet and DNA. The most remarkable new feature is the strong preference for methylation of hemi-methylated over unmethylated DNA displayed by members of the IA and IC family, namely the *EcoKI* and *EcoR124I* systems. The purification of large quantities of the *EcoKI* methyltransferase made more detailed investigations of the protein:protein and protein:DNA interactions possible.

Different methods were used to investigate this preference for hemi-methylated DNA in the *EcoKI* system. A mutational approach identified a number of mutations towards the N-terminus of HsdM which result in enhanced methylation of unmethylated DNA (Kelleher *et al.*, 1991), effectively converting a maintenance methyltransferase into a *de novo* methyltransferase. This clearly implies the N-terminal region of HsdM in the regulation of the substrate specificity, and the region in HsdM harbouring these mutations was termed the  $m^*$  region and the phenotype of these mutants designated  $m^*$ .

The purified *EcoKI* methyltransferase has the expected preference for hemi-methylated DNA (Dryden *et al.*, 1993), but the reason behind this preference remains unknown. An investigation of binding affinities of the methyltransferase complex for differently methylated DNAs found no differences in the binding affinities for un- and hemi-methylated substrates (Powell *et al.*, 1993). Subsequent footprinting experiments however demonstrated different patterns depending on the methylation state of the DNA (see fig. 1o; Powell and Murray, 1995). The conclusion therefore was that whether *EcoKI* restricts or modifies is determined by conformational changes in the enzyme. These experiments also demonstrated the importance of the cofactor





**Fig. 1o: DNA methylation interferes with binding by *M.EcoKI***

Schematic diagram showing the DMS-methylated guanines that interfere with protein binding for (a) unmodified DNA in the absence of AdoMet, or in the presence of AdoHcy, (b) unmodified DNA in the presence of AdoMet or (c) hemi-methylated or fully modified DNA in the presence or absence of AdoMet or AdoHcy. The larger arrows illustrate strong interference at the G indicated, while the smaller ones show weaker interference. The sequence for base pairs 12-38 of the 45mer duplex is shown at the top of the diagram, with the bipartite recognition sequence in bold lettering. The recognition sequence is also indicated on the B-form DNA. Figure adapted from Powell and Murray (1995).



AdoMet for discrimination between unmodified and modified DNA, and showed that this effect could not be mimicked by the presence of the cofactor analogue AdoHcy.

Although it has been suggested that the *EcoKI* enzyme uses the methyl group of bound AdoMet to probe for methyl groups on its substrate DNA, thereby determining the methylation state of the DNA (D.T.F. Dryden, personal communication), it is as yet unclear how the information is communicated from one half-site to the other.

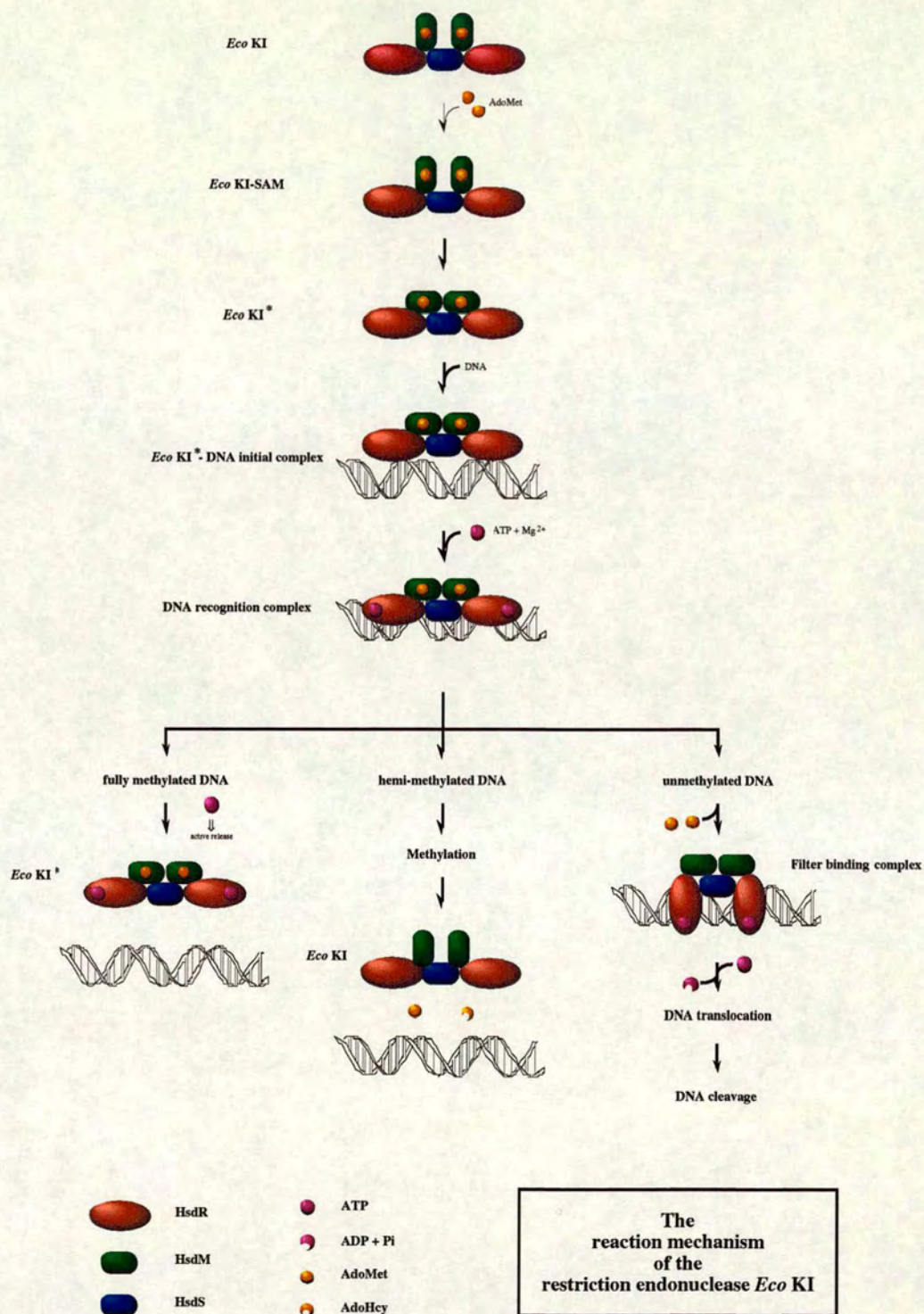
Studies on the *EcoKI* system indicate that the methyltransferase possesses two non-interacting AdoMet binding sites (Powell *et al.*, 1993), indicating that the communication between the two target sites must occur after the binding of AdoMet. Similarly the comparable binding affinities for un- and hemi-methylated DNA make it seem likely that the discrimination event takes place after both substrate and cofactor have been bound, despite the conformational changes observed in *EcoKI* upon binding of both AdoMet and DNA

#### 1.4.3.3. The *EcoKI* system

Most of what is known about the reaction mechanism of the Type I systems *EcoKI* and *EcoBI* has been determined by in vitro studies (Burckhardt *et al.*, 1981a), and a detailed picture of the different steps has been assembled (see fig. 1p).

1. The first step is the rapid binding of AdoMet by *EcoKI*.
2. The *EcoKI*-AdoMet complex then undergoes a slow transition to an activated form *EcoKI*\* (Hadi *et al.*, 1975).
3. *EcoKI*\* forms a non-specific complex with the DNA.
4. If the *EcoKI* recognition site is present on the DNA a more stable complex is formed (Yuan *et al.*, 1975).
5. Depending on the methylation state of the target sequence, different events can be triggered:
  - With a fully methylated site and ATP, the enzyme is released.
  - If the site is hemi-methylated, the unmethylated adenine is rapidly methylated. This occurs some 150 times faster than the methylation of an unmodified target sequence.





**Fig. 1p: The reaction mechanism of the restriction endonuclease EcoK**

Figure from Bickle *et al.* (1978), modified. The arrangement of the R.*EcoKI* subunits relative to each other as well as their contacts with each other as shown in this diagram are purely hypothetical. The dissociation of AdoMet from the filter binding complex could also be due to the loss of one or more subunits from the complex. Although Bickle *et al.* state that "ATP leads to the release of *EcoKI\** from modified DNA", it is not clear in which form the endonuclease dissociates from the DNA. It should also be kept in mind that the final product of the restriction reaction is methylated at the target site (Burckhardt *et al.*, 1981b).



- If the target site is unmodified, ATP induces the transition of *EcoK\** to *EcoK*<sup>+</sup>. This form has lost AdoMet and shows a decrease in diameter, possibly due to rearrangements among the subunits in the complex (Bickle *et al.*, 1978; Burckhardt *et al.*, 1981a). The DNA is then translocated past the enzyme in an ATP-dependent reaction (Rosamond *et al.*, 1979; Yuan *et al.*, 1980), followed by the formation of a single-stranded, then double-stranded break in the DNA at a non-specific site (Meselson and Yuan, 1968; Murray *et al.*, 1973). The enzyme does not turn over in the cleavage reaction but it is still an active ATPase after the cutting of DNA is complete (Eskin and Linn, 1972).
- The final product of the restriction reaction is methylated at the target site.

It is important to note that the *EcoKI* methyltransferase is able to distinguish the different methylation states of the DNA target sequences in the absence of the HsdR subunit. Thus it must be the interaction of HsdM and/or HsdS with the target sequence that determines the enzymatic pathway of the whole system.

#### 1.4.4. Type III enzymes

Among the four Type III systems known to date only the *EcoP1I* and *EcoP15I* systems are well characterised. Genetic analyses show that these systems consist of two genes *res* and *mod* coding for proteins of 106 kDa and 75 kDa respectively. Sequence comparisons reveal that the *res* genes are virtually identical, while the *mod* genes are highly homologous at the 5' and 3' ends but totally unrelated in the central third of the sequence (Iida *et al.*, 1983).

While the Mod subunit alone can catalyse the methylation reaction, both the Res and Mod subunits are necessary for DNA cleavage. The enzymes have an absolute requirement for ATP in the restriction reaction, and cleavage is stimulated by AdoMet, the methyl donor in the modification reaction (Bickle, 1993). Contrary to earlier reports however which state that ATP is not hydrolysed it has now been found that ATP hydrolysis is indeed required for DNA cleavage (Saha and Rao, 1995). In the presence of both ATP and AdoMet restriction and modification are thought to be competing reactions since methylase activity is observed in the course of the



endonucleolytic reaction and only partial digests are obtained (Reiser and Yuan, 1977).

Restriction requires the presence of two inversely oriented unmethylated recognition sites, while methylation proceeds regardless of the orientation of the recognition sequences (Meisel *et al.*, 1992). This means that for cleavage to occur, the true recognition site of Type III restriction enzymes consists of a palindromic sequence interrupted at the centre of symmetry by a spacer of variable length and non-specific sequence. Cleavage is triggered by the close proximity of two convergently tracking EcoP15I-DNA complexes after translocation of DNA in a reaction driven by recognition site-specific ATP hydrolysis (Meisel *et al.*, 1995).

Type III enzymes share many characteristics with Type I enzymes, most notably that the endonuclease and methyltransferase are not independent of each other, and may in fact be regarded as half a Type I enzyme. So while in the Type I enzymes two "half-systems" are connected back to back, therefore translocating DNA in both directions, Type III enzymes can only translocate DNA in one direction. Consequently two Type III enzymes have to come together from opposite directions to achieve the same function as the bi-directional working Type I systems.

### **1.5. Biology and Evolution of R/M-Systems**

The fundamental attribute of a restriction endonuclease is its ability to recognise and selectively attack foreign DNA that enters the cell while its own DNA is protected by the action of a corresponding modification enzyme. Thus it became generally accepted that restriction endonucleases function to prevent the acquisition and expression of foreign DNA, thereby offering a degree of protection against infection by bacteriophages (Arber, 1965b). Although a homogenous population of cells will soon lose any protection due to the emergence of a few modified and therefore resistant phages, a diversity of R/M systems in a population will increase its protection by severely reducing the chances that a phage carrying a specific modification will find and infect a cell with a corresponding R/M system in which it could propagate. Even cells without a R/M system will benefit from this diversity since fewer phage are produced, and will in turn contribute to the restriction of phage by making them susceptible to all the other R/M systems in the population.



The ubiquitous distribution of R/M systems in *enterobacteriae* (Janulaitis *et al.*, 1988; Barcus and Murray, 1995), their extraordinary diversity (Roberts and Macelis, 1996), the discovery of alternate expression of two R/M systems present in the same cell (Glover *et al.*, 1983; Dybvig and Yu, 1994), the selective avoidance of target sites in phage DNA (Sharp, 1986; Krüger *et al.*, 1989; Merkl *et al.*, 1992) and the existence of phage-encoded solitary methyltransferases (Balganesh *et al.*, 1987; Behrens *et al.*, 1987; Trautner *et al.*, 1988; Wilke *et al.*, 1988; Connaughton *et al.*, 1990) and anti-restriction systems (Spoerel *et al.*, 1979; Zabeau *et al.*, 1980; Krüger and Bickle, 1983; Belogurov *et al.*, 1985; Belogurov *et al.*, 1992) all support this view and point to dynamic interactions in the biology and evolution of R/M systems and phage. One might even speculate that the development of a lysogenic cycle in phage is an adaptation to the emergence of R/M systems, since it would enable a phage that was fortunate enough to overcome the restriction barrier to extract maximum benefit from a successful infection by propagating, albeit slowly, in a safe environment.

The complexity of type I and type III systems has led to speculations that R/M systems might have acquired additional functions beside the restriction of phage, and they have been implicated in DNA recombination or repair (Price and Bickle, 1986), regulation of gene expression - which can be influenced by methylation as shown for the *dam* methylase (for a review see Marinus, 1987) - and the acquisition of foreign genes (Roberts, 1978; Roberts, 1978), but there is no conclusive evidence to date to support such alternative functions. Surprisingly however, similarities between the type IIe restriction enzyme *EcoRII* and the integrase family of proteins have been found (Topal and Conrad, 1993), and DNA topoisomerase and recombinase activity has been discovered in a *NaeI* mutant with a single amino acid change (Jo and Topal, 1995), indicating complexities in the type IIe enzymes beyond those needed for restriction.

## 1.6. Aims of this thesis

One of the most puzzling and least understood features of the *EcoKI* R/M system is the *de novo* methylation activity seen in the so called *m\** mutants. However the main obstacle to its investigation were the lack of understanding of how the wild-type system works in the first place, and it was felt that both structural and



biochemical information was needed to address the problem. The main questions were:

- How do the subunits look like (structural information)?
- How do the subunits interact with each other?
- How is the methylation state of the DNA target sequence identified and communicated to the other half site?
- How do cofactor, substrate, and enzyme concentration influence the reaction rate?
- How do AdoMet and DNA binding activities of the mutants compare with the wild-type enzyme?
- How far is information gained from the far simpler Type II R/M systems applicable to the much more complex Type I systems?

In addressing these questions it has become possible to propose a new model for the structure and activity of the *EcoKI* methyltransferase which is consistent with earlier results. The model also explains how the m\* mutants affect the methylation activity of M.*EcoKI* on unmethylated DNA (see chapter 6).



## **CHAPTER 2:**

# **MATERIALS & METHODS**



## 2.1. Bacterial Strains

**Table 2.1 Bacterial Strains**

Strain	relevant features	Reference / Source
BL21 ( $\lambda$ DE3)	$r_B^- m_B^- sup^0 hsdS_B$ , ( $\lambda$ DE3)	(Studier and Moffat, 1986)
C600	$r_K^+ m_K^+$	(Appleyard, 1954)
NM522	$hsd \Delta 5$ ( $r_K^- m_K^-$ ), $lacI^q lacZ \Delta M15$	(Gough and Murray, 1983)
NM679	$\Delta$ ( $mrr-hsd-mcr$ )	(King and Murray, 1995)
NM679 (DE3)	NM679 lysogenic for $\lambda$ DE3	(King and Murray, 1995)
$\lambda$ DE3	Bacteriophage $\lambda$ carrying the gene for T7 RNA polymerase	
$lacI^q$	Overproduction of the <i>lac</i> repressor protein leads to high levels of the <i>lac</i> repressor protein	
$lacZ \Delta M15$	Partial deletion of $\beta$ -D-galactosidase gene, allows complementation of $\beta$ -galactosidase activity by $\alpha$ -complementation sequence (present in certain cloning vectors) and therefore blue-white screen for recombinant colonies when plated on X-Gal.	
$r_K^+$ or $r_K^-$	restriction proficient or deficient for <i>Eco</i> KI	
$m_K^+$ or $m_K^-$	modification proficient or deficient for <i>Eco</i> KI	



## 2.2. Phage Strains

**Table 2.2 Phage Strains**

Strain	Relevant Features	Reference / Source
NM 1048	$\lambda$ <i>hsdMS</i> <sup>+</sup> , cI857	(Sain and Murray, 1980)
NM 1090	$\lambda$ ra118cI60	lab collection
NM 1094	$\lambda$ ra118cI60	lab collection
$\lambda$ 1305-7	$r^{+m*}$ Glu48 $\rightarrow$ Lys	(Kelleher et al., 1991)
$\lambda$ 1305-27	$r^{+m*}$ Arg72 $\rightarrow$ Gln	(Kelleher et al., 1991)
$\lambda$ 1305-5, -10	$r^{+m*}$ Leu85 $\rightarrow$ Gln	(Kelleher et al., 1991)
$\lambda$ 1305-8, -17	$r^{-m*}$ Leu113 $\rightarrow$ Arg	(Kelleher et al., 1991)
$\lambda$ 1048-13	$r^{-m*}$ Leu113 $\rightarrow$ Pro	(Kelleher et al., 1991)
$\lambda$ 1305-19, -20 -21, -22 -23, -24	$r^{-m*}$ Leu113 $\rightarrow$ Gln	(Kelleher et al., 1991)
$\lambda$ 1048-16	$r^{-m*}$ Trp115 $\rightarrow$ Arg	(Kelleher et al., 1991)
$\lambda$ 1305-6, -28	$r^{+m*}$ Arg124 $\rightarrow$ Leu	(Kelleher et al., 1991)
$\lambda$ 1305-18	$r^{+m*}$ Arg124 $\rightarrow$ Cys	(Kelleher et al., 1991)
$\lambda$ 1305-11	$r^{+m*}$ Leu134 $\rightarrow$ Val	(Kelleher et al., 1991)
$\lambda$ 1305-25	$r^{-m*}$ Ser144 $\rightarrow$ Tyr	(Kelleher et al., 1991)
$\lambda$ 1305-9	$r^{+m*}$ Arg153 $\rightarrow$ His	(Kelleher et al., 1991)
$\lambda$ 1305-12	$r^{+m*}$ Arg153 $\rightarrow$ Ser	(Kelleher et al., 1991)
$\lambda$ 1305-26	$r^{+m*}$ Arg153 $\rightarrow$ Cys	(Kelleher et al., 1991)



## 2.3. Plasmids

**Table 2.3 Plasmids**

Plasmid	description	Relevant Features	Reference / Source
pJF118EH	overexpression vector	tac promoter	(Fürste <i>et al.</i> , 1986)
pJF118HE	overexpression vector	tac promoter	(Fürste <i>et al.</i> , 1986)
pJF-E48K	overexpresses hsdM <sub>E48K</sub> S	hsdMS <sub>K</sub> in pJF118	This work
pJF-R72Q	overexpresses hsdM <sub>R72Q</sub> S	hsdMS <sub>K</sub> in pJF118	This work
pJF-L85Q	overexpresses hsdM <sub>L85Q</sub> S	hsdMS <sub>K</sub> in pJF118	This work
pJF-L113R	overexpresses hsdM <sub>L113R</sub> S	hsdMS <sub>K</sub> in pJF118	This work
pJF-L113P	overexpresses hsdM <sub>L113P</sub> S	hsdMS <sub>K</sub> in pJF118	This work
pJF-L113Q	overexpresses hsdM <sub>L113Q</sub> S	hsdMS <sub>K</sub> in pJF118	This work
pJF-W115R	overexpresses hsdM <sub>W115R</sub> S	hsdMS <sub>K</sub> in pJF118	This work
pJF-R124L	overexpresses hsdM <sub>R124L</sub> S	hsdMS <sub>K</sub> in pJF118	This work
pJF-R124C	overexpresses hsdM <sub>R124C</sub> S	hsdMS <sub>K</sub> in pJF118	This work
pJF-L134V	overexpresses hsdM <sub>L134V</sub> S	hsdMS <sub>K</sub> in pJF118	This work
pJF-S144Y	overexpresses hsdM <sub>S144Y</sub> S	hsdMS <sub>K</sub> in pJF118	This work
pJF-R153H	overexpresses hsdM <sub>R153H</sub> S	hsdMS <sub>K</sub> in pJF118	This work
pJF-R153S	overexpresses hsdM <sub>R153S</sub> S	hsdMS <sub>K</sub> in pJF118	This work
pJF-R153C	overexpresses hsdM <sub>R153C</sub> S	hsdMS <sub>K</sub> in pJF118	This work
pT7-5	overexpression vector	T7 gene 10 promoter	(Tabor, 1990)
pT7-7	overexpression vector	T7 gene 10 promoter, RBS, ATG start cocon	(Tabor, 1990)



## 2.4. Oligonucleotides

All oligonucleotides were made by and purchased from Oswel DNA Service's branch at the University of Edinburgh (except for the T7-7 primer N6728 which was kindly provided by Rob van Nues). The information from Oswel's data sheets supplied with each oligonucleotide was used for subsequent calculations.

**Table 2.4 Oligonucleotides**

Name	Sequence
408	5'-TGTCTAGATATCGGCCT <b>AACC</b> ACGTGG <b>TGCG</b> TACGAGCTCAGGCG-3'
409	3'-ACAGATCTATAGCCGGAT <b>TTGG</b> TGCACC <b>CACG</b> CATGCTCGAGTCCGC-5'
860	5'-TGTCTAGATATCGGCCT <b>A6CC</b> ACGTGG <b>TGCG</b> TACGAGCTCAGGCG-3'
861	3'-ACAGATCTATAGCCGGAT <b>TTGG</b> TGCACC <b>6CGC</b> CATGCTCGAGTCCGC-5'
M3105	5'-CGGCCT <b>AACC</b> ACGTGG <b>TGCG</b> TACGA-3'
M3106	3'-GCCGGAT <b>TTGG</b> TGCACC <b>CACG</b> CATGCT-5'
N8624	5'-CGGCCT <b>A6CC</b> ACGTGG <b>TGCG</b> TACGA-3'
N8625	3'-GCCGGAT <b>TTGG</b> TGCACC <b>6CGC</b> CATGCT-5'
P8870	5'-CGGCCT <b>AACC</b> ACGTGGTGC <b>TGCG</b> TACGA-3'
P8871	3'-GCCGGAT <b>TTGG</b> TGCACC <b>6CGC</b> CATGCT-5'
P8872	5'-CGGCCT <b>A6CC</b> ACGTGG <b>TGCG</b> TACGA-3'
P8873	3'-GCCGGAT <b>TTGG</b> TGCACC <b>6CGC</b> CATGCT-5'
G3270	5'-CTAGGA <b>AACT</b> CCATGG <b>TGCG</b> GACG-3'
G3271	3'- <b>CTTG</b> AGGTACC <b>CACG</b> CTGCAGCT-5'
N6728	5'-AATACGACTCACTATAG-3'
N8111	5'-GCAAGCTTAGTACTGGCCTGCACCAGACTAGGT-3'

6 = 6-methyl-adenine



## 2.5. Enzymes and Chemicals

Restriction enzymes were supplied by Bethesda Research Laboratories (UK) Ltd, Boehringer Mannheim (UK) Ltd, New England Biolabs, or Northumbria Biologicals Ltd.

Low melting point agarose, ethidium bromide, bromophenol blue, and Ficoll were purchased from Sigma Chemical Company. Agarose was supplied by Miles Laboratory Ltd and Flowgen Instruments Limited. GeneClean II™ kit was from Bio 101 Incorporated, California, USA.

T4 DNA ligase, dNTPs and ddNTPs were from Pharmacia; Sequenase enzyme from USB; DNA polymerase from New England Biolabs (NEB); DNase I, RNase A and lysozyme from Sigma Chemical Company Ltd.; T4 polynucleotide kinase (T4 PNK) from S. Bruce (Institute of Cell and Molecular Biology, University of Edinburgh).

Standard laboratory chemicals were supplied by BDH Chemicals Ltd, Fisons Scientific Equipment, or Sigma Chemical Company unless otherwise stated.

Ampicillin (Penbritin) was from Beecham Pharmaceuticals; DTT,  $\beta$ -mercaptoethanol and IPTG from Sigma; X-gal from Boehringer.

Nitrocellulose filters were from Schleicher and Schuell; Digoxigenin labelling kit and anti-digoxigenin-AP Fab fragments from Boehringer; AMPPD from Tropix (purchased through New Brunswick Scientific); autoradiographic film from Amersham International.

Radiolabelled compounds ( $[\alpha\text{-}^{32}\text{P}]\text{dCTP}$ ,  $[\alpha\text{-}^{35}\text{S}]\text{dATP}$ ,  $[\gamma\text{-}^{32}\text{P}]\text{ATP}$  and  $[\text{methyl-}^3\text{H}]\text{AdoMet}$ ) were from Amersham International. Scintillation fluid from National Diagnostics. Acrylamide and *N,N*-methylene bis-acrylamide were from Bethesda Research Laboratories (UK) Ltd and Northumbria Biologicals Ltd.

AdoMet was a gift from New England Biolabs or from Sigma. HPLC purified ANS from Molecular probes. Coomassie blue R250, MES, tris (hydroxymethyl) amino-methane (Tris) and SDS were from Sigma. Spectroscopy was done on a Perkin Elmer Lambda 5 spectrophotometer. Chromatography media and columns were from Pharmacia, and HPLC was done on a system from Gilson Medical Electronics Inc.



## 2.6. Media and materials

All media were sterilised by autoclaving (15 lb in<sup>-2</sup>, 15 minutes).

L-Broth:	Difco Bacto tryptone (10 g), Difco Bacto yeast extract (5 g), NaCl (5 g), dH <sub>2</sub> O to 1 litre; adjusted to pH 7.2 with NaOH before autoclaving.
L-agar:	Difco Bacto tryptone (10 g), Difco Bacto yeast extract (5 g), NaCl (5 g), Difco agar (15 g), dH <sub>2</sub> O to 1 litre; adjusted to pH 7.2 with NaOH before autoclaving.
BBL-Agar:	Baltimore Biological Labs. Trypticase (10 g), NaCl (5 g), Difco agar (10 g), dH <sub>2</sub> O to 1 litre.
BBL-Top Agar:	As for BBL-agar but only 6.5g Difco agar added per litre.
2x TY Broth:	Difco Bacto tryptone (16 g), Difco Bacto yeast extract (10 g), NaCl (10 g), dH <sub>2</sub> O to 1 litre.
5x Spizen Salts:	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> (10 g), K <sub>2</sub> HPO <sub>4</sub> (70 g), KH <sub>2</sub> PO <sub>4</sub> (30 g), tri-sodium citrate dihydrate (30 g), MgSO <sub>4</sub> •7H <sub>2</sub> O (1 g), dH <sub>2</sub> O to 1 litre.
Phage Buffer:	Na <sub>2</sub> HPO <sub>4</sub> (7 g), KH <sub>2</sub> PO <sub>4</sub> (3 g), NaCl (5 g), MgSO <sub>4</sub> •7H <sub>2</sub> O (10 ml, 0.01 M), CaCl <sub>2</sub> (10 ml, 0.01 M), gelatine (1 ml, 1 % w/v), dH <sub>2</sub> O to 1 litre.
X-gal Indicator plates:	X-gal (20 µl, 20 mg ml <sup>-1</sup> ), IPTG (20 µl, 20 mg ml <sup>-1</sup> ), per 2.5 ml of BBL top agar.
Dialysis tubing:	Dialysis tubing was prepared by cutting dry tubing into 20 cm strips, boiling it for 10 minutes (in a large volume 2 % w/v sodium bicarbonate, 1 mM EDTA, pH 8), rinsing it with dH <sub>2</sub> O, and boiling it again for 10 minutes (in 1 mM EDTA, pH 8). The tubing was finally stored in 50 % v/v ethanol, 1 mM EDTA at 4 °C, and rinsed with dH <sub>2</sub> O before use.



## 2.7. Standard Solutions

### 2.7.1. Standard solutions used for microbial techniques

SOC:	L-Broth supplemented with 0.36 % Glucose, 10 mM $\text{MgSO}_4$ , 10 mM $\text{MgCl}_2$
Ampicillin:	100 mg $\text{ml}^{-1}$ stock solution in $\text{dH}_2\text{O}$ , 100 $\mu\text{g ml}^{-1}$ end concentration. Stored at 4 °C.
Chloramphenicol:	34 mg $\text{ml}^{-1}$ stock solution in ethanol, 25 $\mu\text{g ml}^{-1}$ end concentration. Stored at -20 °C.
IPTG:	20 mg $\text{ml}^{-1}$ (84 mM) stock solution in $\text{dH}_2\text{O}$ , 1 mM end concentration. Stored at -20 °C.

### 2.7.2. Standard solutions used for DNA preparations

Lysis solution:	Tris-HCl (25 mM, pH 8.0), EDTA (50 mM), stored at 4 °C
Alkaline SDS:	NaOH (0.2 N), SDS (1 % w/v), stored at RT
Neutralisation solution:	KAc (3 M), Acetic Acid (2 M), stored at 4 °C
LiCl:	5 M, stored at -20 °C.
NaCl / PEG:	NaCl (1.6 M), Polyethylene glycol (PEG 8000, 13 %)
P / C / I:	Phenol / Chloroform / Isoamylalcohol (25:24:1)
TE Buffer:	Tris (10 mM), EDTA (1 mM), adjusted to appropriate pH with HCl.

### 2.7.3. Standard solutions used for Hybridization

20x SSC:	NaCl (3 M), tri-sodium citrate (0.3 M), pH 7.0.
----------	---



### **2.7.3.1. for Colony Lifts**

Lysis solution:	10 % SDS
Denaturing solution:	NaOH (0.5M), NaCl (1.5 M)
Neutralizing solution:	Tris (0.5 M, pH 7.5), NaCl (1.5 M)
Washing solution:	2x SSC

### **2.7.3.2. for Hybridization**

Blocking solution:	10 % w/v Blocking reagent (Boehringer), autoclaved and stored at 4 °C
Hybridization buffer:	Formamide (50 %), 5x SSC, Blocking reagent (2 %), SDS (0.02 % w/v), N-laurolyl-sarcosine (0.1 % w/v)
Washing solution 1:	SSC (2x), SDS (0.1 %)
Washing solution 2:	SSC (1x), SDS (0.1 %)

### **2.7.3.3. for Digoxigenin detection**

MS Buffer:	Maleic Acid (0.1 M), NaCl (0.15 M), pH to 7.5 with NaOH; made up as 5x solution and autoclaved.
Washing buffer:	MS buffer with 0.3 % v/v Tween 20
Blocking buffer:	MS buffer with 1 % Blocking reagent
TMS buffer:	Tris (0.1 M, pH 9.5), NaCl (0.1 M), MgCl <sub>2</sub> (50 mM).
AMPPD:	[3-(2'-spiroadamantane)-4-methoxy-4-(3'-phosphoryloxy)phenyl-1,2-dioxethane], the chemiluminescence substrate for alkaline phosphatase was supplied as a 10 mg ml <sup>-1</sup> solution (from Tropix, purchased through New Brunswick Scientific). A 1:100 diluted solution in TMS buffer was stored at 4 °C and could be reused 10 times.



#### **2.7.4. Standard solutions used for DNA agarose gel electrophoreses**

Ethidium Bromide: 10 mg ml<sup>-1</sup> in dH<sub>2</sub>O. Stored in dark at 4 °C.

50x TAE: Tris (2 M), Acetic Acid (5.71 % v/v), EDTA (50 mM)

20x TBE Buffer: Tris (1.78 M), boric acid (1.78 M), EDTA (50 mM).

6x Gel-loading-Buffer: Bromophenol blue (0.25 %), xylene cyanol FF (0.25 %), Ficoll (15 %), stored at RT

#### **2.7.5. Standard solutions used during protein purification**

1x TM buffer: 20 mM Tris, 20 mM MES, 10 mM MgCl<sub>2</sub>, 0.1 mM EDTA, 7 mM β-Mercaptoethanol; pH 6.5 or 8.0

PMSF: 20 mM stock solution in ethanol. Stored in dark at 4 °C. Used at 10 nM end concentration.

Benzamidine: 10 mM stock solution in dH<sub>2</sub>O. Stored in dark at 4 °C. Used at 10 nM end concentration.

#### **2.7.6. Standard solutions used for PAGE**

Acrylamide stock: 30 % w/v acrylamide stock solution (acrylamide:bisacrylamide in a ratio of 37.5:1).

APS: 10 % w/v in dH<sub>2</sub>O

##### **2.7.6.1. Standard solutions used for non-denaturing PAGE**

5 % Acrylamide: 5 % Acrylamide:bisacrylamide in 1x TBE

2x Loading buffer: Tris-HCl (100 mM, pH 8.0), glycerol (20 %)

Running buffer: 1x TBE



#### **2.7.6.2. Standard solutions used for SDS-PAGE**

Stacking gel mix:	Acrylamide:bisacrylamide (5 % w/v), Tris-HCl (125 mM, pH 6.8), SDS (0.1 %)
Separating gel mix:	Acrylamide:bisacrylamide (12 % w/v), Tris-HCl (375 mM, pH 8.8), SDS (0.1 %)
SDS Loading buffer:	Tris-HCl (125 mM, pH 6.8), glycerol (10 %), SDS (2 %), $\beta$ -mercaptoethanol (0.7 M), bromophenol blue (2.5 mg ml <sup>-1</sup> ).
SDS running buffer:	Tris-HCl (25 mM), glycine (192 mM), SDS (0.1 %)

#### **2.7.7. Standard solutions used for gel retardation**

1x Binding buffer:	Tris-HCl (20 mM, pH 8.0), NaCl (100 mM), Glycerol (5 %)
1x Dissociation buffer:	Tris-HCl (20 mM, pH 8.0), NaCl (100 mM), MgCl <sub>2</sub> (7 mM), $\beta$ -mercaptoethanol (10 mM), Glycerol (5 %)

#### **2.7.8. Standard solutions used for ANS assay**

ANS assay buffer:	Tris-HCl (20 mM, pH 8.0), NaCl (100 mM), MgCl <sub>2</sub> (7 mM), $\beta$ -mercaptoethanol (10 mM), Glycerol (5 %)
-------------------	---

#### **2.7.9. Standard solutions used for HPLC filtration**

HPLC buffer:	Tris-HCl (20 mM, pH 6.5), NaCl (200 mM), MgCl <sub>2</sub> (10 mM), $\beta$ -mercaptoethanol (7 mM), EDTA (0.1 mM)
--------------	--



#### **2.7.10. Standard solutions used for Methyltransferase assay**

Methylation buffer: Tris-HCl (100 mM, pH 8.0), NaCl (100 mM), MgCl<sub>2</sub> (7 mM), β-mercaptoethanol (10 mM)

STOP solution: Ammonium hydrogencarbonate (NH<sub>4</sub>HCO<sub>3</sub>, 0.2 M)



## **2.8. Microbial Techniques**

### **2.8.1. Growth of *E. coli* bacterial cultures**

- 1) L-Broth (with or without antibiotic as required) was inoculated from a single colony with a sterilised inoculating loop or a sterile toothpick.
- 2) Cultures were grown at 37 °C with shaking in either flasks or glass bottles until the desired density was reached.

### **2.8.2. Storage of bacterial strains**

- 1) Bacterial strains were stored by adding 0.5 ml of a fresh overnight culture to 0.5 ml cell storage buffer, frozen on liquid nitrogen, and stored at -70 °C.
- 2) To recover the cells the surface of the frozen culture was scratched with a sterilised inoculating loop and streaked out on a plate containing the appropriate antibiotic.

### **2.8.3. Preparation of plating cells and bacterial lawns**

- 1) A fresh overnight culture was diluted 20fold in L-Broth (with antibiotic if required) and grown to mid log-phase at 37 °C.
- 2) The cells were harvested by centrifugation (3000 g, 5 minutes), resuspended in half the original volume of MgSO<sub>4</sub> (10 mM), and stored on ice until used.
- 3) A lawn of cells was made by putting 0.2 ml of the plating cells into a sterile 5 ml tube, adding 3 ml of molten top agar, inverting the tube twice, and pouring the mixture onto an agar plate (containing the appropriate antibiotic if required).



#### **2.8.4. Preparation of $\lambda$ phage plate lysates**

- 1) Phage suspension (0.1 ml, containing approximately  $10^5$  phage) and fresh plating cells (0.2 ml) were mixed in a sterile 5 ml tube, and the phage left to adsorb for 15 minutes.
- 2) Molten top agar (3 ml) was added and the mixture poured onto a fresh L-agar plate (containing antibiotic if necessary).
- 3) The plate was incubated (not inverted) for 6 to 8 hours at 37 °C until confluent lysis was observed.
- 4) L-broth (5 ml) was added and the plate incubated at 4 °C overnight.
- 5) The L-broth was decanted into a bottle, 100  $\mu$ l chloroform were added, and the lysate vortexed.
- 6) Cellular debris was removed by centrifugation (3000 g, 10 minutes) and the supernatant constituting the phage lysate transferred to a fresh bottle and stored at 4 °C.

#### **2.8.5. Phage Titration**

##### **2.8.5.1. Spot Tests**

- 1) Serial dilutions of phage stock were prepared using phage buffer.
- 2) Plating cells (0.2 ml) were spread onto a BBL plate (containing the appropriate antibiotic if required) in BBL top agar (3 ml).
- 3) 10  $\mu$ l of the serial dilution of  $\lambda$  phage were spotted onto the bacterial lawn and allowed to dry before overnight incubation at 37 °C.

##### **2.8.5.2. Plate titrations**

- 1) Serial dilutions of phage stock were prepared using phage buffer.



- 2) 0.1 ml of each dilution was added to plating cells (0.2 ml) in a sterile 5 ml tube, vortexed briefly, and left to adsorb for 15 minutes.
- 3) The mixture was plated out onto BBL plates (containing the appropriate antibiotic if required) in BBL-top agar (3 ml) and incubated overnight at 37 °C.

#### **2.8.6. Preparation and Transformation of Competent Cells**

The preparation of competent cells for transformation by the  $\text{CaCl}_2$  method was superseded when the equipment for the more efficient electroporation method became available. The DNA being used for electroporation was additionally purified using the GeneClean II™ kit (see 2.9.4.4.) to remove salt before being used for transformation.

##### **2.8.6.1. $\text{CaCl}_2$ method**

- 1) A fresh overnight culture was diluted 100fold in L-broth and grown at 37 °C to an O.D.<sub>650nm</sub> of 0.5.
- 2) The cells were harvested by centrifugation (3000 g, 5 minutes).
- 3) The cells were resuspended in an equal volume of ice-cold  $\text{MgCl}_2$  (100 mM).
- 4) The cells were harvested again (3,500 g, 5 minutes) and resuspended in one half the original volume of ice-cold  $\text{MgCl}_2$  (100 mM).
- 5) The cells were harvested for a third time (4000 g, 5 minutes) and resuspended in one tenth the original volume of ice-cold  $\text{CaCl}_2$  (100 mM).
- 6) 200 µl of the competent cells were mixed in an eppendorf with 1-5 µl of a ligation reaction or about 100 ng of purified plasmid DNA and left on ice for 30 minutes.
- 7) The mixture was heat-shocked in a 42 °C water bath for exactly 90 seconds before being returned onto ice for a further 30 minutes.
- 8) 1 ml L-Broth was added and the cells incubated for 1 hour in a 37 °C water bath before 50-200 µl of the cells were spread on agar plates containing the appropriate antibiotic and incubated at 37 °C overnight.



#### **2.8.6.2. Electroporation**

- 1) A fresh overnight culture was diluted 100fold into 10 ml 2x TY and grown at 37 °C to an O.D.<sub>650nm</sub> of 0.5.
- 2) The cells were chilled on ice for 20 minutes before being harvested by centrifugation (3000 g, 5 minutes).
- 3) The cells were resuspended in 10 ml of ice-cold dH<sub>2</sub>O and harvested three times (at 3000 g, 3500 g, 4000 g respectively for 5 minutes each).
- 4) After the third time the cells were resuspended in 0.5 ml of ice-cold dH<sub>2</sub>O and kept on ice until used.
- 5) 40 µl of the competent cells were mixed in an eppendorf with 0.5-1 µl of a ligation reaction or about 10 ng of purified plasmid DNA and kept on ice.
- 6) The mixture was transferred to a pre-cooled electroporation cuvette using sterile pasteur pipettes and electroporated (at 2.5 kV, 200 Ω resistance and 25 µF capacitance) with a Bio-rad Gene Pulser™/Pulse Controller.
- 7) Immediately after electroporation 1 ml SOC buffer was added into the cuvette and the cells transferred to a sterile 25 ml Universal.
- 8) After incubation for 1 hour at 37 °C between 50 and 200 µl of the cell suspension was spread on agar plates containing the appropriate antibiotic and incubated at 37 °C overnight.

#### **2.8.7. Detection of positive Transformants**

##### **2.8.7.1. Colony Lifts**

- 1) The required number of Whatman No. 1 paper was cut down to an appropriate size, and Hybond™-N circular nitrocellulose filters were labelled with a soft pencil and cut marks to facilitate the correct alignment of filters and plates for the subsequent identification of positive transformants.



- 2) Plates were taken from the incubator when the colonies were about 1-2 mm in diameter and stored at 4 °C for about 5 minutes to ensure reliable transfer of colonies.
- 3) Four pieces of Whatman No. 1 paper were placed onto large glass plates and the required solutions - 10 % SDS (Lysis solution), 0.5 M NaOH / 1.5 M NaCl (Denaturing solution), 0.5 M Tris / 1.5 M NaCl, pH 7.5 (Neutralizing solution), and 2x SSC (Washing solution) respectively - poured onto them. The solutions were evenly distributed by rolling a 10 ml pipette along the glass plate so that the paper was saturated with solution but not soaking wet.
- 4) The labelled nitrocellulose filter discs were laid onto the surface of the agar plates and left for 1 minute to allow colony transfer. After marking the plates to allow later orientation with respect to the discs, the filters were lifted from the plates.
- 5) The plates were returned to the incubator and kept at 37 °C for about 6 hours to allow regrowth of the colonies before storage at 4 °C.
- 6) The filters were placed colony side uppermost onto the prepared moist Whatman No. 1 paper for 5 minutes each in the order:
  - a) Lysis solution (10 % SDS)
  - b) Denaturing solution (0.5 M NaOH / 1.5 M NaCl)
  - c) Neutralizing solution (0.5 M Tris / 1.5 M NaCl, pH 7.5)
  - d) Washing solution (2x SSC)

In between transfers the filters were briefly put onto a dry piece of blotting paper to remove excess liquid.

- 7) The filters were finally blotted dry and the DNA fixed to the filter by ultra-violet irradiation in a Stratagene UV Stratalinker® 1800 set to auto-crosslink (15 min at 120,000  $\mu\text{joules cm}^{-2}$ ).

#### **2.8.7.2. Hybridization**

- 1) Bacterial debris was gently scraped from the filters using cotton wool soaked in 2x SSC, and 7.5 ml Hybridization buffer (50 % formamide, 5x SSC, 2 %



Blocking solution, 0.02 % SDS, 0.1 % w/v N-laurolyl-sarcosine) were prepared per filter.

- 2) The filters were sealed in polythene bags with 6 ml Hybridization buffer per filter and prehybridized for 1 hour at 37 °C with gentle shaking.
- 3) The digoxigenin-labelled DNA probe (see 2.9.3.6.) was added to the remaining Hybridization buffer ( $\approx 10 - 20 \text{ ng ml}^{-1}$  end concentration), denatured by boiling for 10 minutes and immediately placed on ice to prevent renaturation.
- 4) The prehybridization solution was discarded and replaced by the hybridization solution containing the digoxigenin-labelled DNA probe, and hybridization continued overnight at 37 °C with gentle shaking.
- 5) The hybridization solution was recovered from the polythene bags and stored at -20 °C for reuse, and the filters washed twice in 100 ml 2x SSC, 0.1 % SDS for 5 minutes each before being washed twice in 100 ml 1x SSC, 0.1 % SDS for 15 minutes each.

#### **2.8.7.3. Digoxigenin detection**

The digoxigenin detection method was carried out at room temperature, unless otherwise stated, with filters incubated on a gently shaking platform.

- 1) The filters were washed for 5 minutes in 100 ml Washing buffer (1x MS buffer, 0.3 % v/v Tween 20), then placed in 100 ml Blocking buffer (1x MS buffer, 1 % w/v Blocking reagent) for 30 minutes.
- 2) Binding of anti-digoxigenin-AP Fab fragments (3  $\mu\text{l}$  anti-DIG-AP in 30 ml Blocking buffer,  $\approx 75 \text{ mUnits per ml}$ ) took place for 30 minutes before the filters were washed twice in 100 ml Washing buffer (1x MS buffer, 0.3 % v/v Tween 20) for 15 minutes each and once in 50 ml TMS buffer (0.1 M Tris (pH 9.5), 0.1 M NaCl, 50 mM  $\text{MgCl}_2$ ) for 5 minutes.
- 3) The filters were incubated for 5 minutes in 20 ml of a 1:100 diluted solution of the chemiluminescence reagent AMPPD in TMS buffer before being blotted for a few seconds and sealed in Saran wrap together with a few drops of TMS buffer.



- 4) The saran wrap containing the filters was put into an open film cassette and exposed to light at 37 °C for 15 minutes before exposure to X-ray film for a maximum of 30 minutes (5 minutes often proved to be enough).
- 5) Positive transformants were identified and picked after aligning the X-ray autograph with the corresponding plate.

## **2.9. DNA Techniques**

### **2.9.1. Basic Techniques**

#### **2.9.1.1. Phenol / Chloroform extraction**

This method was used to remove proteins (*e.g.* during plasmid preparations, after digestion of DNA with restriction enzymes, after ligations) from a DNA containing solution. Since very high DNA concentrations (*e.g.* after large scale plasmid preparations) can result in a phase inversion between the DNA containing aqueous and the protein containing organic phase the identity of each phase was confirmed by either its colour or its smell.

- 1) An equal volume of TE saturated phenol was added to an eppendorf containing the DNA solution and thoroughly mixed. The organic and aqueous layers were separated by centrifugation (11,000 g, 5 min) and the aqueous layer containing the DNA transferred to a new eppendorf.
- 2) An equal volume of TE saturated phenol / Chloroform / Isoamylalcohol (25:24:1) was added to the DNA solution and again thoroughly mixed. The organic and aqueous layers were separated by centrifugation (11,000 g, 5 min) and the aqueous upper layer transferred to a new eppendorf.
- 3) To remove all traces of phenol which might interfere with subsequent enzymatic reactions an equal volume of Chloroform / Isoamylalcohol (24:1) was added to the eppendorf containing the DNA solution. After thorough mixing the organic and aqueous layers were again separated by centrifugation (11,000 g, 5 min) and the aqueous layer containing the DNA transferred to a new eppendorf.



### **2.9.1.2. Precipitation of DNA with Ethanol and Sodium Acetate**

- 1) DNA in solution was precipitated by adding 0.1x volume of sodium acetate (3 M, pH 5.0) and 2.5x volume of ethanol. The mixture was then incubated at -70 °C for 30 minutes.
- 2) The precipitate was harvested (11,000 g, 10 minutes) and twice washed with ethanol (70 % w/v).
- 3) The pellet was dried under vacuum and resuspend in the required volume of TE buffer (pH 8.0).

### **2.9.2. Preparation of plasmid DNA**

#### **2.9.2.1. Alkaline Mini-Prep**

This method is based upon that of Birnboim and Doly (1979) and Ish-Horowicz and Burke (1981).

- 1) A fresh overnight culture (1.5 ml) carrying the plasmid was harvested by centrifugation (11,000 g, 5 minutes).
- 2) The pellet was resuspended in 100 µl ice-cold lysis solution (Tris-HCl, 25 mM; EDTA, 10 mM; glucose, 0.9 %) and incubated on ice for 5 minutes.
- 3) 200 µl alkaline SDS (NaOH, 0.2 M; SDS, 1 %) was added and mixed by gentle inversion, followed by incubation on ice for 5 minutes.
- 4) 150 µl ice-cold potassium acetate (3 M, pH 4.8) were added, followed by incubation on ice for 5 minutes.
- 5) The supernatant was harvested by centrifugation (11,000 g, 15 minutes) and transferred into a new eppendorf cap.
- 6) Ethanol (2.5x volume, -20 °C) was added to the supernatant to precipitate the plasmid DNA and, after mixing by inverting the eppendorf caps at least three times, incubated at -70 °C for 30 minutes .



- 7) The precipitated plasmid DNA was harvested by centrifugation (11,000 g, 15 minutes) and washed with ethanol (70 %,w/v).
- 8) The DNA pellet was dried under vacuum and resuspended in TE buffer (20 µl, pH 8.0).

#### **2.9.2.2. Boiling Method**

This method is based upon that of Holmes & Quigley (1981).

1. A fresh overnight culture (1.5 ml) carrying the plasmid was harvested by centrifugation (11,000 g, 5 minutes).
2. The supernatant was removed as completely as possible and the pellet resuspended in 200 µl lysis solution (Tris-HCl, pH 8.0, 50 mM; EDTA, pH 8.0, 50 mM; Triton X-100, 0.5 % (v/v); sucrose, 8 %).
3. The solution was boiled for exactly 40 seconds, then cooled on ice for 10 min before the removal of the cells debris by centrifugation (11,000 g, 15 min).
4. The “snotty” pellet was removed using a yellow tip, 20 µl 3 M NaAc and 200 µl Isopropanol were added and the solution incubated at -70 °C for 15 minutes.
5. The precipitated plasmid DNA was harvested by centrifugation (11,000 g, 10 minutes), the supernatant removed completely, and the pellet rinsed with ether.
6. After the ether evaporated completely the pellet was resuspend in 50 µl TE (pH 8.0).

#### **2.9.2.3. QIAprep™ Plasmid Preparation**

This method is based upon the modified alkaline lysis method of Birnboim and Doly (1979) and on the adsorption of DNA onto silica in the presence of high salt (Vogelstein and Gillespie, 1979). It is fast and produces very high quality DNA, but was only seldom used in order to cut costs. The procedure was followed as indicated by the manufacturer.



#### 2.9.2.4. Large scale preparation of plasmid DNA

This protocol is based on an unpublished procedure by R. Treisman as described in "Molecular Cloning" (Sambrook *et al.*, 1989), and is fast with a very high yield of high quality plasmid.

##### *Amplification of the plasmid*

- 1) 200 ml LB (containing the appropriate antibiotic) were inoculated with 2 ml cell culture and grown to an OD<sub>600nm</sub> of approximately 1.2.
- 2) 120 µl 30 mg ml<sup>-1</sup> chloramphenicol (end concentration 50 µg ml<sup>-1</sup>) were added and the culture grown o/n for 14 h.

##### *Harvest and Lysis of Bacteria*

- 1) The cells were harvested by centrifugation (Sorvall, GS-3-rotor, 7000 rpm, 10 min, 4 °C), resuspended in 20 ml lysis solution (50 mM EDTA, 25 mM Tris-HCl, pH 8.0) to dissolve the cell walls and put on ice for 5 min.
- 2) While carefully shaking the suspension 40 ml "alkaline SDS" (0.2 N NaOH, 1 % SDS) were slowly added and the mixture put on ice for 10 min to permit the dissolution of the cells membranes and precipitation of the cells DNA, RNA and protein contents.
- 3) 30 ml "neutralisation solution" (3 M KAc, 2 M Acetic Acid) were added while gently (in order to avoid breaking the chromosomal DNA) swirling the viscous solution, and the mixture left on ice for another 5 min to allow low molecular weight DNA and RNA to renature and go back into solution.
- 4) The mixture was spun (Sorvall, GS-3-rotor, 10,000 rpm, 15 min, 4 °C, without brakes) to remove the precipitate and the supernatant filtered through a ready-made funnel (stuffed with glass wool, washed first with ethanol, then dH<sub>2</sub>O) into a clean centrifuge bottle.
- 5) 0.6 Vol. Isopropanol (≈ 54 ml) were added to the cleared supernatant, mixed well, and left at RT for 10 min.



- 6) The solution was spun down (Sorvall, GS-3-rotor, 10,000 rpm, 15 min, at RT), the supernatant discarded, and the bottle left standing in an inverted position until the last traces of isopropanol evaporated.
- 7) The pellet was dissolved in TE (pH 8.0) and the volume upped to a total of 3 ml.

*Purification of plasmid-DNA by precipitation with Polyethylene glycol*

- 1) The DNA solution was transferred into a 15 ml corex tube and 3 ml cold 5 M LiCl (stored at -20 °C) added to precipitate high molecular weight RNAs.
- 2) The precipitate was removed by centrifugation (Sorvall, SS34-rotor, 10,000 rpm, 10 min, 4 °C) and the supernatant transferred to a fresh 30 ml corex tube.
- 3) 1 Vol. Isopropanol ( $\approx$  6 ml) were added, the solution left at RT for 10 min, and the precipitate harvested by centrifugation (Sorvall, SS34-rotor, 10,000 rpm, 10 min, at RT)
- 4) The supernatant was discarded and the corex tube left standing in an inverted position until the last traces of isopropanol evaporated.
- 5) The pellet was dissolved in 500  $\mu$ l TE (pH 8.0) containing RNase (20  $\mu$ g ml<sup>-1</sup> end conc.) and the solution transferred into an eppendorf cap and incubated at RT for 30 min.
- 6) The plasmid DNA was precipitated through the addition of 500  $\mu$ l 1.6 M NaCl / 13 % Polyethylene glycol (PEG 8000) and harvested by high speed centrifugation (microfuge, 12,000 rpm, 5 min, 4 °C)
- 7) The supernatant was transferred into a new eppendorf and set aside, while the pellet containing the plasmid DNA was resuspended in 400  $\mu$ l TE (pH 8.0).
- 8) 400  $\mu$ l Phenol/Chloroform/Isoamylalcohol (25:24:1) were added, the solution vortexed for 1 min, spun down for 15 sec, and the upper aqueous phase transferred into a new eppendorf cap.
- 9) The Phenol/Chloroform/Isoamylalcohol-extraction was repeated and followed by a Chloroform/Isoamylalcohol (24:1) extraction to remove all traces of phenol. Since very high DNA concentrations can result in a phase inversion between the



DNA containing aqueous and the organic phase the identity of each phase was confirmed by its smell.

- 10) The plasmid DNA was precipitated from the aqueous phase by the addition of 0.1 Vol. 3 M potassium acetate and 2.5 Vol. ethanol (mixing was insured by inversion of the eppendorf cap for at least three times), followed by incubation at -70 °C for 30 min.
- 11) The precipitated plasmid was harvested by centrifugation (Microfuge, 11,000 rpm, 30 min, 4 °C) and washed twice with 1 ml 70 % ice cold ethanol (stored at -20 °C).
- 12) The supernatant was removed and discarded as completely as possible and the pellet dried under vacuum (5 min) before being resuspended in 100 µl TE (pH 8.0).
- 13) The DNA-concentration was measured by scanning a 1:200 dilution from 200 nm to 300 nm and taking readings at 260 nm and 280 nm. The  $A_{260\text{nm}}$  value was used to calculate the DNA concentration while the  $A_{260\text{nm}} / A_{280\text{nm}}$  ratio was used to give an indication of the purity of the preparation (pure DNA has an  $A_{260\text{nm}} / A_{280\text{nm}}$  ratio of  $\approx 1.8$ ).

### **2.9.3. DNA modifications**

#### **2.9.3.1. Restriction Endonuclease Digestion of DNA**

Restriction endonuclease digestion of DNA was carried out in volumes of 10-50 µl containing 0.2-2.0 µg DNA using the buffers supplied and under conditions specified by the suppliers for the respective restriction enzyme. Double digests were performed in buffers and under conditions suitable for both restriction enzymes unless no such conditions could be found, in which case a short purification step using the GeneClean II™ kit was included before the second digest.

#### **2.9.3.2. Phosphorylation of DNA fragments**

If unphosphorylated linear DNA fragments (*e.g.* synthetic oligonucleotides or PCR products) were to be used in a ligation reaction, they were first phosphorylated



by incubation with 5 units T4 polynucleotide kinase and 0.1 mM ATP in 1x PNK buffer (50 mM Tris-HCl, pH 7.5; 10 mM MgCl<sub>2</sub>; 5 mM DTT) at 37 °C for 45 minutes and subsequently purified using the GeneClean II™ kit (see 2.9.4.4.).

#### **2.9.3.3. Creating "blunt-ended" DNA fragments**

If necessary 5'-overhangs in DNA fragments (*e.g.* in restriction products) were filled in using 1 unit Klenow polymerase and 33 μM dNTPs in 1x buffer (50 mM Tris-HCl, pH 7.2; 10 mM MgSO<sub>4</sub>; 0.1 mM DTT). After incubation of the mixture for 30 minutes at room temperature the DNA was purified using the GeneClean II™ kit (see 2.9.4.4.).

#### **2.9.3.4. Ligation of DNA**

Ligation of DNA using T4 DNA ligase was carried out in volumes of 10-20 μl containing 0.2-0.4 μg DNA in 1x buffer (30 mM Tris-HCl, pH 7.8; 10 mM MgCl<sub>2</sub>; 10 mM DTT; 500 μM ATP) provided by the supplier. Incubation was at 16 °C overnight and was terminated by heat inactivation at 70 °C for 10 minutes.

#### **2.9.3.5. End Labelling of Oligonucleotides**

Oligonucleotides are synthesized without a phosphate group at their 5' termini and were therefore easily labelled by phosphorylation using the reaction described in 2.9.3.2. and substituting ATP for [γ-<sup>32</sup>P]ATP. The reaction was terminated by incubation of the mixture at 68 °C for 10 minutes, and unincorporated [γ-<sup>32</sup>P]ATP was removed using a 1 ml G25 spun column spun for 3 min in a bench top centrifuge.

#### **2.9.3.6. Preparation of digoxigenin DNA probes**

Digoxigenin DNA probes were prepared using the Boehringer Mannheim kit, and the procedure followed as indicated by the manufacturer.

- 1) 0.5 to 3 μg of DNA was denatured for 10 minutes in a boiling water bath and placed on ice for 3 minutes before adjusting the volume to 14 μl.
- 2) 2 μl hexanucleotide mix in 10x reaction buffer (vial 5), 2 μl dNTP labelling mix (vial 6), and 2 μl 1U μl<sup>-1</sup> Klenow enzyme were added and the mixture incubated at 37 °C overnight.



- 3) The reaction was stopped with the addition of 2 µl 0.2 M EDTA, and the DNA precipitated with 1.8 µl 3 M NaAc and 72 µl ethanol, left at -20 °C for 30 minutes, and harvested by centrifugation (10 minutes at 11,000 g).
- 4) The pellet was washed with 70 % ethanol, dried, and resuspended in 50 µl TE and 0.5 µl 10 % SDS.
- 5) The labelled DNA was stored at -20 °C after incubation at 37 °C for 10 minutes with occasional vortexing.
- 6) The success of the labelling reaction was tested by spotting dilutions of the probe and of control DNA supplied with the kit onto Hybond™-N hybridization transfer membrane, cross-linking the DNA to the membrane with ultra-violet irradiation and developing as described in 2.8.7.3. (Digoxigenin detection). A comparison between the quantified control and the probe allowed an estimate of the concentration of labelled DNA.

#### **2.9.3.7. Annealing of complementary oligonucleotides**

Equimolar amounts of two complementary oligonucleotides were combined and heated to 90-100 °C for 5 min, followed by stepwise cooling for 10 minutes at 68 °C, 42 °C, 37 °C, and on ice respectively before storage at -20 °C.

#### **2.9.3.8. Site-directed mutagenesis by PCR**

The mutagenesis method employed used a primer for the PCR reaction designed to contain a mismatch within the sequence complementary to the template (in order to introduce the desired mutation) as well as a sequence at the 5' end with a suitable restriction site not present in the template (to allow selective cloning of the PCR products).

PCR reactions were carried out using a Hybaid Omni-Gene thermal cycler, and the annealing temperature was calculated according to the equation:

$$T_{\text{Annealing}} = T_m - 5^{\circ}\text{C} = 4 * (\sum G + \sum C) + 2 * (\sum A + \sum T) - 5 [^{\circ}\text{C}]$$

where  $\sum G$ ,  $\sum C$ ,  $\sum A$ , and  $\sum T$  are the numbers of G, C, A, and T nucleotides respectively in the oligonucleotide. Since small changes in the annealing temperature were found to have a profound influence on the PCR reaction, three reactions were



performed in parallel at various temperatures close to the calculated annealing temperature in order to find the optimum conditions for a specific pair of primers.

- 1) 100 µl 1x reaction buffer (10 mM KCl, 10 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 20 mM Tris-HCl (pH 8.8), 2 mM MgSO<sub>4</sub>, 0.1 % Triton X-100) containing 0.4 µM of both upstream and downstream primer, 250 µM of each dNTP, and 2 U Vent<sub>R</sub> DNA polymerase were added to about 60 pg template DNA in a 0.5 ml eppendorf tube and overlaid with 70 µl mineral oil.
- 2) The thermal cycle started with an initial denaturation step (95 °C for 2 minutes), was followed by 28 amplification cycles (denaturation at 95 °C for 1 minute, annealing at the calculated temperature for 30 seconds, extension at 72 °C for 1 minute), and finished with a final extension step (72 °C for 5 minutes).
- 3) The PCR reaction products were separated on a TAE agarose gel, stained with ethidium bromide, visualised under long-wave UV light, and the required DNA fragment cut out and purified using the GeneClean II™ method (see 2.9.4.4.).
- 4) The purified DNA fragment was digested using the appropriate restriction enzymes, purified using the GeneClean II™ method, and cloned into a suitable vector for further analysis.

#### **2.9.4. Analysis of DNA fragments on and purification from agarose gels**

##### **2.9.4.1. Agarose Gel Electrophoresis**

For both analysis of DNA on standard agarose and purification of DNA on low melting-point agarose, horizontal submerged gel electrophoresis equipment was used. Agarose gels were made in the concentration range 0.7-1.5 % w/v, depending on anticipated DNA fragment size. A 1x TAE buffer system was used if the DNA was to be purified from the agarose by the GeneClean II™ method, otherwise an 1x TBE buffer system was given preference. DNA samples (0.2-0.5 µg) and marker DNA of known size were applied in combination with Ficoll loading dye (20 % w/v in dH<sub>2</sub>O, with bromophenol blue) and electrophoresis carried out at 10 V cm<sup>-1</sup>. Following staining with ethidium bromide, DNA was visualised under long-wave UV light and,



if necessary, the section containing the desired DNA fragment cut out from the agarose gel.

#### **2.9.4.2. DNA Fragment Preparation by Electroelution**

- 1) The gel section containing the desired DNA fragment was placed in dialysis tubing together with an equal volume of 1x TBE buffer, the tubing submerged in 1x TBE buffer, and a potential difference of 10 V cm<sup>-1</sup> applied.
- 2) The gel slice was removed after 1 hour and checked under long-wave UV light to ensure that the DNA had eluted before the current was reversed for 10 seconds to move the DNA from the side of the tubing.
- 3) The TBE buffer containing the DNA was transferred to a 1.5 ml eppendorf tube, and the DNA purified by phenol/chloroform extraction followed by ethanol precipitation.
- 4) The resulting DNA pellet was resuspended in a suitable volume of TE (typically 20 µl).

#### **2.9.4.3. DNA Fragment Preparation from Low Melting Point Agarose Gels**

- 1) The DNA band cut from the gel was added to 1x volume TBE, NaCl (0.2 M) in a 1.5 ml eppendorf tube followed by melting in a 65 °C water bath.
- 2) Protein was removed by extracting twice with phenol (pre-equilibrated with 1x TBE, 0.1 M NaCl), and ethidium bromide was removed by extracting the aqueous layer thrice with 4x volume butanol.
- 3) The DNA was precipitated with ethanol and sodium acetate, and the resulting pellet resuspended in a suitable volume of TE (typically 20 µl).

#### **2.9.4.4. DNA Fragment Preparation with GeneClean II™**

This method, which was found to be both fast and reliable, was used to purify DNA fragments from agarose gels, to remove contaminating proteins from a DNA



containing solution, and to change buffer conditions from high to low salt (*e.g.* after ligations, restriction digests).

- 1) The gel section or solution containing the desired DNA fragment was placed in a 1.5 ml eppendorf tube together with 3 volumes 3 M NaI.
- 2) If the DNA was purified from a gel then the agarose was melted at 68 °C for about 10 minutes before the addition of 5 µl glassmilk per 5 µg DNA.
- 3) The resulting suspension was left standing at room temperature for 10 minutes to allow adsorption of the DNA to the glassmilk with occasional inversion of the eppendorf tube.
- 4) The suspension was spun down (Microfuge, 11,000 rpm, 5 seconds) and the supernatant removed as completely as possible.
- 5) Thrice the pellet was taken up in 400 µl "New Wash" solution (supplied by the manufacturer), precipitated again by centrifugation (Microfuge, 11,000 rpm, 5 seconds) and the supernatant discarded.
- 6) After the final step the pellet was allowed to air-dry for 5 minutes before it was resuspended in 10 µl dH<sub>2</sub>O and incubated for 3 minutes at 55 °C to facilitate the desorption of the DNA from the silicate particles of the glassmilk..
- 7) The glassmilk was spun down (Microfuge, 11,000 rpm, 30 seconds) and the supernatant containing the DNA transferred to a new eppendorf tube.



## **2.10. Protein Techniques**

### **2.10.1. Expression of hsdMS**

Expression of plasmids encoding mutant methyltransferases was tested by induction of a 10 ml culture with 1 mM IPTG for 6 hours. 1 ml of the culture was transferred to an eppendorf tube, spun down, resuspended in 100  $\mu$ l TM buffer (20 mM Tris, 20 mM MES, 10 mM  $\text{MgCl}_2$ , 0.1 mM EDTA, 7 mM  $\beta$ -Mercaptoethanol; pH 6.5 or 8.0) supplemented with 20  $\mu$ M benzamidine and 10  $\mu$ M PMSF to inhibit proteases, and sonicated for 30 seconds. 20  $\mu$ l of the cell extract were transferred to a fresh eppendorf tube while the remainder was spun for 15 minutes at 4 °C in a microfuge at maximum speed. 20  $\mu$ l of the resulting supernatant were transferred to another fresh eppendorf tube, and an equal volume SDS sample buffer added to both cell extract and supernatant. 5  $\mu$ l of each sample were run and analysed on a SDS-polyacrylamide gel.

### **2.10.2. Protein Purification and Storage**

The procedure was based on the protocol established by Dryden *et al.* (1993), and the same TM buffer (20 mM Tris, 20 mM MES, 10 mM  $\text{MgCl}_2$ , 0.1 mM EDTA, 7 mM  $\beta$ -Mercaptoethanol; pH 6.5 or 8.0), supplemented with 20  $\mu$ M benzamidine and 10  $\mu$ M phenylmethylsulphonyl fluoride to inhibit proteases, was used throughout the purification procedure. All chromatography columns were run in the cold room at 48 ml hour<sup>-1</sup> unless otherwise stated, and 5 ml fractions collected every 6.25 minutes. 5  $\mu$ l samples were taken from the fractions and analysed on SDS-PAGE for the presence of methyltransferase, and those fractions containing methyltransferase were pooled and dialysed for at least 4 hours against 2 l TM buffer to remove salt and adjust the pH before being applied onto the next column. Columns were washed with at least 2 volumes 1 M NaCl in TM buffer after use and re-equilibrated in TM buffer of the appropriate pH.

- 1) Using 2 l baffled flasks to increase aeration, freshly transformed NM679 cells were grown in 500 ml L-Broth (with 50  $\mu$ M Ampicillin) to an O.D.<sub>650</sub> of  $\approx$  0.5. Expression was induced by adding IPTG to 1 mM and growth was continued for



6 hours. Cells were then harvested by centrifugation (4000 g, 15 minutes, 4 °C) and the cell paste (typically about 9 g) stored at -70 °C until required. All subsequent steps in the purification were performed at 4 °C.

- 2) The cell paste was resuspended in 5 ml TM buffer (pH 6.5) per gram cell paste, put in an ice-water bath, and sonicated in 30 second bursts for a total time in minutes equal to the weight of the cell paste in gram.
- 3) The resulting cell-free extract was clarified by centrifugation (30,000 g, 3 hours), and the supernatant applied to a CM-Sepharose column (33 x 2.6 cm) pre-equilibrated in TM buffer (pH 6.5). The column was washed with at least one column volume of buffer to remove unbound material, and the bound protein eluted with a 500 ml gradient of 0 - 0.5 M NaCl in TM buffer (pH 6.5).
- 4) The dialysed fractions were applied to a Heparin-Agarose column (15 x 1.6 cm) pre-equilibrated in TM buffer (pH 8.0). The column was washed with at least one column volume of buffer to remove unbound material, and the bound protein eluted with a 500 ml gradient of 0 - 1.0 M NaCl in TM buffer (pH 8.0).
- 5) The dialysed fractions were applied to a DEAE-Sepharose column (20 x 1.4 cm) pre-equilibrated in TM buffer (pH 8.0). The column was washed with at least one column volume of buffer to remove unbound material, and the bound protein eluted with a 500 ml gradient of 0 - 0.5 M NaCl in TM buffer (pH 8.0). The two peaks containing methyltransferase subunits were pooled separately, precipitated by adding ammonium sulphate to 80 % saturation, and the pellets obtained after centrifugation resuspended in 1 ml TM buffer (pH 6.5).
- 6) The resuspended protein solutions were further purified on a gel filtration column (20 x 1.4 cm), pre-equilibrated in 200 mM NaCl in TM buffer (pH 6.5) and run at 24 ml hour<sup>-1</sup>, 5 min frac<sup>-1</sup>, 2 ml frac<sup>-1</sup>. The originally used S300 gel filtration column was later replaced by a Hi-Load 16/60 Superdex 200 gel filtration column which has a much better resolution. The final pooled fractions were precipitated by adding ammonium sulphate to 80 % saturation, and the pellets obtained after centrifugation resuspended in storage buffer (1x TM, 200 mM NaCl, 50 % Glycerol) and stored at -20 °C.



### **2.10.3. Polyacrylamide Gel Electrophoresis (PAGE)**

All PAGE was carried out with either large (13.5 cm by 13.5 cm by 1 or 2 mm) Atto AE-6220 gel kits (Atto Corporation, Japan) or Hoefer "Mighty Small" (7.5 cm by 8 cm by 0.75 mm) SE250 gel kits (Hoefer Scientific Instruments, San Francisco, USA). Polymerisation of the respective gel mixes was started by the addition of APS and TEMED to an end concentration of 0.1 %.

#### **2.10.3.1. Non-denaturing PAGE**

Non-denaturing PAGE was used to separate labelled DNA fragments and for gel retardation experiments. Samples were mixed with an equal volume of 2x Loading buffer before loading, and the polyacrylamide gels run at 30 mA, with 5 µl 0.25 % Xylene cyanol (migrates at  $\approx$  260 bp) / 0.25 % Bromophenol blue (migrates at  $\approx$  65 bp) in separate lanes at the edge of the gel to monitor the progress of the samples, until the desired separation was achieved.

#### **2.10.3.2. SDS-PAGE**

SDS-PAGE in a Tris-glycine buffer system (Laemmli, 1970) was used to separate and visualise proteins on polyacrylamide gels. Samples were boiled for 3 minutes with an equal volume SDS loading buffer before being loaded onto a 5 % stacking gel / 12 % solving gel and run in SDS running buffer at 30 mA until the bromophenol blue dye had reached the bottom of the gel.

#### **2.10.3.3. Staining of proteins with Coomassie**

To visualise proteins after their separation by SDS PAGE the gels were stained in an aqueous solution of ethanol (250 ml l<sup>-1</sup>), formaldehyde (10 ml l<sup>-1</sup> of a 35 % stock solution) and Coomassie Brilliant Blue R-250 (1.2 g l<sup>-1</sup>) for about 30 minutes. The stain was removed by immersion of the gel in a destaining solution of 10 % v/v methanol / 10 % v/v acetic acid placed in a 60 °C water bath amid gentle shaking, and periodically changing the destain solution.



#### 2.10.4. Desalting of protein

Glycerol and salt were removed and the storage buffer exchanged against the respective assay buffer by running 20 - 50 µl protein stock solution through a G-25 Sephadex gel filtration column (PD10 from Pharmacia) pre-equilibrated in the assay buffer. 10 drop fractions were collected, the fractions containing the protein pooled, and the protein concentration determined.

#### 2.10.5. Determination of protein concentration in solution

Most proteins have a distinct absorption maximum at 280 nm, due primarily to the presence of tyrosine, tryptophan, and phenylalanine, and a mixture of proteins in solution at a concentration of 1 mg ml<sup>-1</sup> has an average extinction coefficient  $\epsilon_{280\text{nm}}$  of about 1 M<sup>-1</sup> cm<sup>-1</sup>. Partially purified preparations (*e.g.* cell extracts) may also contain nucleic acids that have an absorption maximum at 260 nm. The concentration of proteins in solution can therefore be estimated from measurements of the absorption of UV-light at 260 and 280 nm according to the following equation (Harris, 1987):

$$\text{protein [mg ml}^{-1}\text{]} \approx 1.55 * A_{280\text{nm}} - 0.76 * A_{260\text{nm}}$$

However the extinction coefficient of purified proteins can deviate considerably from this average value, and since their specific molar extinction coefficient can be estimated from the sum of values of the absorbance of tyrosine, tryptophane and phenylalanine residues in a protein of known amino acid sequence (Mihalyi, 1970), the following equation was used to calculate the concentration of purified proteins:

$$A_{280\text{nm}}: \sum \text{Phe} * 0.7 + \sum \text{Trp} * 5559 + \sum \text{Tyr} * 1197 \cong 1 \text{ M}$$

Protein ( <i>Eco</i> KI)	MW [D]	$A_{280\text{nm}}$ of an 1 M solution	$A_{280\text{nm}}$ of an 1 mg ml <sup>-1</sup> solution	concentration [nM] of a 1 mg ml <sup>-1</sup> solution
M	59 313.51	58850.7	0.992	16860
S	51 401.91	26267.7	0.511	19454
R	124 834.28	114199.8	0.915	8011
M <sub>2</sub>	118 627.02	117701.4	0.992	8430
M <sub>1</sub> S <sub>1</sub>	110 715.42	85118.4	0.769	9032
M <sub>2</sub> S <sub>1</sub>	170 028.93	143969.1	0.847	5881



The absorption spectrum between 240 and 340 nm was routinely used to examine purity and aggregation of proteins in solution, and a peak-to-trough ratio ( $A_{280\text{nm}} : A_{250\text{nm}}$ ) of 2.5 - 2.8 was usual.

## **2.11. Assays**

### **2.11.1. Gel retardation assays**

Gel retardation assays use the fact that DNA-protein complexes move slower through a non-denaturing polyacrylamide gel than free DNA, thereby allowing the separation of the two forms. By using labelled DNA the two forms can be visualised and quantified by autoradiography.  $K_d$  values for binding can then be estimated by determination of the protein concentration needed to complex half of the DNA under conditions where the protein was in excess over the DNA (Hendrickson and Schleif, 1984).

Signals were quantified by densitometry using a Shimadzu dual wavelength thin layer chromatoscanner (model CS-930) densitometer. Depletion of free DNA provides a more accurate measure of complex formation than bound DNA because free DNA cannot change once it has entered the gel. Bound DNA, however, can dissociate within the gel resulting in a smearing effect below the complex, making quantification more difficult (Fried, 1989).

#### **2.11.1.1. DNA binding assay**

DNA binding experiments were essentially done as described by Powell *et al.* (1993). In order to estimate relative dissociation constants ( $K_d$  values), a constant amount of DNA was titrated with increasing amounts of methyltransferase, and the complexes formed separated from free DNA on 5 % polyacrylamide gels.

- 1) Methyltransferase in 1x binding buffer (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 5 % Glycerol) was prepared at a range of concentrations (0 - 200 nM in the absence, 0 - 20 nM in the presence of 100  $\mu\text{M}$  AdoMet), and 10  $\mu\text{l}$  of each concentration distributed into eppendorf tubes and kept on ice.



- 2) 0.2 nM  $^{32}\text{P}$ -labelled oligonucleotide in 1x binding buffer ( $\pm$  100  $\mu\text{M}$  AdoMet as required) was prepared, 10  $\mu\text{l}$  added to each eppendorf tube containing the methyltransferase, and the solution carefully mixed by stirring.
- 3) After incubation on ice for 10 minutes to allow the formation of DNA-protein complexes the samples were loaded onto 5 % non-denaturing polyacrylamide gels and run at 30 mA for 1 hour.
- 4) The gels were dried at 80  $^{\circ}\text{C}$  for 1.5 hours and autoradiographed using preflashed film.

#### **2.11.1.2. Dissociation assay**

A constant amount of methyltransferase bound to  $^{32}\text{P}$ -labelled DNA was titrated with an excess of unlabelled DNA for increasing amounts of time, and the complexes separated from free DNA on 5 % polyacrylamide gels.

- 1) A mixture of 1  $\mu\text{M}$  methyltransferase and 200 nM  $^{32}\text{P}$ -labelled oligonucleotide in 1x dissociation buffer (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 7 mM  $\text{MgCl}_2$ , 10 mM  $\beta$ -mercaptoethanol, 5 % Glycerol) was prepared ( $\pm$  100  $\mu\text{M}$  AdoMet), incubated for 10 minutes at 4  $^{\circ}\text{C}$  to allow formation of the protein-DNA complexes, and 10  $\mu\text{l}$  distributed into eppendorf tubes and kept on ice.
- 2) 10  $\mu\text{l}$  20  $\mu\text{M}$  unlabelled oligonucleotide in 1x binding buffer ( $\pm$  100  $\mu\text{M}$  AdoMet as required) were added to each eppendorf tube at various times, resulting in a 100fold excess of unlabelled DNA, and the solution carefully mixed by stirring.
- 3) At the end of the incubation period all samples were loaded onto 5 % non-denaturing polyacrylamide gels and run at 30 mA for 1.5 hours.
- 4) The gels were dried at 80  $^{\circ}\text{C}$  for 1.5 hours and autoradiographed using preflashed film.

#### **2.11.2. AdoMet Binding by Fluorescence of the extrinsic Fluorophore ANS**

A methyltransferase / ANS solution was titrated with increasing amounts of AdoMet, and the displacement of protein bound ANS by AdoMet monitored by



fluorescence spectroscopy. All fluorescence spectra and titrations were measured in a Perkin Elmer LS50 fluorimeter using a 400  $\mu$ l quartz microcuvette with a 2 mm pathlength. Experiments were performed at 25 °C and temperature was controlled to  $\pm 0.1$  °C.

- 1) Stock solutions of 1.05  $\mu$ M methyltransferase, 3 mM AdoMet ( $A_{260\text{nm}}$  of a 1M solution = 15400), and 1 mM ANS ( $A_{350\text{nm}}$  of a 1M solution = 4950) were prepared in 1x ANS assay buffer (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 7 mM  $\text{MgCl}_2$ , 10 mM  $\beta$ -mercaptoethanol, 5 % Glycerol).
- 2) Fluorescence spectra were taken for 1x ANS assay buffer, 50  $\mu$ M ANS in 1x ANS assay buffer, 1.05  $\mu$ M methyltransferase in 1x ANS assay buffer, and 1  $\mu$ M methyltransferase / 50  $\mu$ M ANS in 1x ANS assay buffer. Excitation was at 295 and 395 nm, and emission was measured from 400 to 600 nm with a 10 nm slit width.
- 3) 400  $\mu$ l solution containing 1  $\mu$ M methyltransferase and 50  $\mu$ M ANS in 1x ANS assay buffer were titrated with AdoMet over the concentration range 0 - 300  $\mu$ M. The solution was left standing for 1 minute in the fluorimeter after the addition of AdoMet to allow equilibration, and 10 measurements were taken at each AdoMet concentration. Excitation was at 395 nm, emission was at 480 nm, bandwidths were 10 nm, and the integration time was set to 5 seconds.
- 4) Data were corrected for the inner filter effect and dilution, and analysed using the Grafit data analysis programme (Leatherbarrow, 1992).

### 2.11.3. HPLC filtration

The state of the methyltransferase complex at different protein concentrations was investigated by HPLC. Serial dilutions of the  $\text{M}_2\text{S}_1$  complex in HPLC buffer (20 mM Tris-HCl, pH 6.5; 200 mM NaCl; 10 mM  $\text{MgCl}_2$ ; 7 mM  $\beta$ -mercaptoethanol; 0.1 mM EDTA) ranging from 2400 nM to 40 nM were applied to a 4.6 x 250 mm Hydropore-5-SEC column preceded by a guard column, and the elution volume determined for each concentration. A reference peak, which is probably due to traces of glycerol left in solution after desalting of the methyltransferase on a PD10 column, served as internal standard.



#### 2.11.4. Methyltransferase assays

The enzyme catalysed transfer of [*methyl* -<sup>3</sup>H] from [*methyl* -<sup>3</sup>H]AdoMet onto DNA was followed with the analysis of aliquots taken from the reaction at various time points.

- 1) Microtiter plates were labelled and 200 µl STOP solution (0.2 M NH<sub>4</sub>HCO<sub>3</sub>) pipetted into each well.
- 2) DNA, AdoMet, and methyltransferase stock solutions of the appropriate concentration were made up in Methylation buffer (100 mM Tris-HCl, pH 8.0, 100 mM NaCl, 7 mM MgCl<sub>2</sub>, 10 mM β-mercaptoethanol), aliquoted where necessary, and put on ice until use.
- 3) Reactions were set up by adding 12 µl DNA and 8 µl AdoMet to labelled eppendorf tubes, and the mixtures and the separate methyltransferase solution incubated at 25 °C for exactly 10 minutes.
- 4) Reactions were initiated (time 0) by the addition of 5 µl methyltransferase solution to 20 µl DNA/AdoMet mixture, and the solution mixed by swirling of the pipette tip.
- 5) 2 µl aliquots were withdrawn 15 seconds before selected timepoints during the reaction, and transferred into the appropriate microtiter wells containing the STOP solution at precisely the selected timepoints.
- 6) The samples were dot-blotted onto DE81 paper previously wetted with H<sub>2</sub>O, and washed once with 1 ml STOP solution and twice with 1 ml ethanol.
- 7) The DE81 paper was dried for at least 30 min at 80 °C or at 65 °C o/n, cut into appropriate pieces, and distributed into scintillation vials.
- 8) 2 ml EcoScint were added to the samples, and the <sup>3</sup>H count determined in a Beckman LS 6500 scintillation counter.



## 2.12. Software used

Wordprocessing:	Microsoft Word 5.1, Microsoft Word 6.0, WriteNow 3.0
Bibliography:	EndNote Plus 2.1
Graphics:	MacDraw, MacDraw Pro, ClarisDraw
Sequence analysis:	GCG7, GeneJockey II, Lasergene DNA*
Data analysis:	Excel 5, GraFit 3.0
WorldWideWeb browser:	Netscape Navigator 2.02
Molecular modelling:	FRODO, SYBIL (by Tripos)



## **CHAPTER 3:**

# **CLONING AND EXPRESSION**



### 3.1. Introduction

A total of 14 mutants (summarized in table 2.2) expressing an unusual *de novo* methylation activity had previously been isolated in our laboratory (Kelleher *et al.*, 1991). Sequencing had revealed that the mutations are located in the N-terminal third of the *hsdM* gene, and the relevant region was dubbed the m\* region and the phenotype of the mutants called m\*. The biochemical characterisation of these mutants is dependent on the availability of large amounts of the mutant proteins, and since these m\* mutants had been isolated as  $\lambda$  phages carrying the *hsdMS* genes, it was necessary to transfer the *hsdMS* genes into a suitable vector for over-expression of the genes and purification of the mutant enzymes. The plasmid pJF118HE was chosen because it had proven to be an effective system for the expression and purification of the wildtype *EcoKI* methyltransferase (Dryden *et al.*, 1993).

Several attempts at crystallisation of the *EcoKI* methyltransferases had so far been unsuccessful (D.T.F. Dryden, personal communications). The construction of C-terminal deletion derivatives of the HsdM subunit was attempted on the hypothesis that the N-terminal m\* region might encode a separate domain (see chapter 5). Consequently it might be possible to express, purify, and crystallise the proposed domain on its own. A suitable *ScaI* restriction site located at the end of the m\* region in the *hsdM* gene was chosen to construct a precise C-terminal deletion derivative of HsdM.

The mechanism by which the m\* mutants affect the methylation activity of the *EcoKI* methyltransferase is unknown. Evidently information about the methylation state at one half site of the bipartite recognition sequence has to be relayed to the other half site, and the clustering of the m\* mutations in the N-terminal third of *hsdM* implicate this region in the relay function. Two principle mechanisms are conceivable. In the first one the proposed m\* domain could block methylation at the other half site unless one half site is already methylated and the block removed, while in the second one the proposed m\* domain would activate methylation at the other half site if the first half site is already methylated. Consequently mutations that inactivate the block in the first mechanism or lead to a continuous activation of the methyltransferase in the second would both display *de novo* methylation activity. However deletion of the proposed m\* domain would have very different effects depending on which



mechanism applies. In the "Block" model the deletion should result in a continuously active methyltransferase, comparable in its activity to Type II methyltransferases, while the "Activation" model would predict the deletion to totally inactivate the methyltransferase. N-terminal deletion derivatives of the HsdM subunit were constructed using the same *ScaI* restriction site as for C-terminal deletion derivatives, and the truncated *hsdMS* genes transferred into suitable expression vectors.

### 3.2. Subcloning of m\* mutants

#### 3.2.1. Subcloning of *hsdMS* from $\lambda$ phage into the expression vectors pJF118HE or pJF118EH

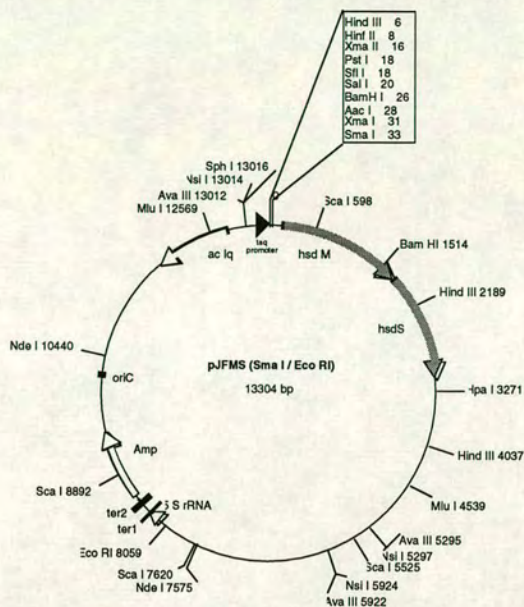
For the expression of the wildtype methyltransferase an 8026 bp *SmaI-EcoRI* DNA fragment (comprising the full *hsdM* and *hsdS* genes as well as 120 bp upstream of *hsdM* and 4925 bp downstream of *hsdS*) had been inserted into the 5281 bp plasmid pJF118HE cut with the same restriction enzymes. The resulting 13304 bp plasmid was designated pJFMS<sub>*SmaI/EcoRI*</sub> (see Fig. 3.1.a), and expressed the wildtype methyltransferase from an IPTG induceable *tac* promoter (Dryden *et al.*, 1993).

The transfer of the *hsdMS* genes harbouring the m\* mutation from  $\lambda$  phage into pJF118HE by this route proved unexpectedly difficult, and only three of the mutants (Leu113 → Gln, Leu134 → Val, Arg153 → His) were transferred as *SmaI-EcoRI* DNA fragments. The likely reason for the difficulties seemed to be the close proximity of the *SmaI* and *EcoRI* restriction sites in the polylinker region of pJF118HE.

An alternative strategy used a suitable *HpaI* restriction site located 587 bp downstream of *hsdS* to transfer the remaining mutations as an 3688 bp *SmaI-HpaI* DNA fragment into pJF118HE (cut with *SmaI* and dephosphorylated to prevent religation of the plasmid). If the DNA fragment inserted the right way round, then the resulting 8968 bp plasmid was designated pJFMS<sub>*SmaI/HpaI*</sub>, otherwise pJFMS<sub>*HpaI/SmaI*</sub> (see Fig. 3.1.b). In cases where the fragment inserted the wrong way round, a 3700 bp *SmaI-SalI* DNA fragment containing the *hsdMS* genes was excised and inserted into pJF118EH (which differs from pJF118HE in the orientation of the polylinker site), resulting in a 8958 bp plasmid designated pJFMS<sub>*SmaI/SalI*</sub> (Fig. 3.1.c).



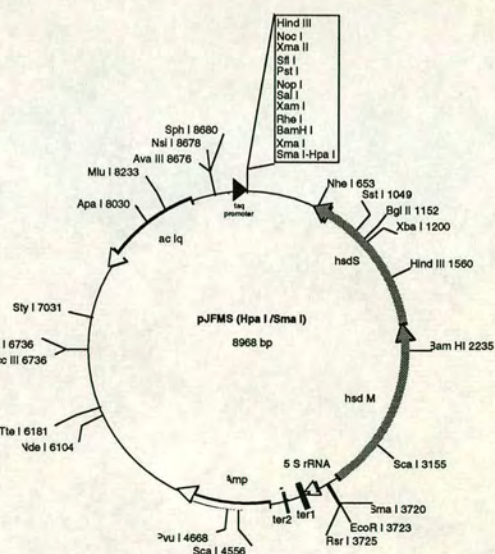
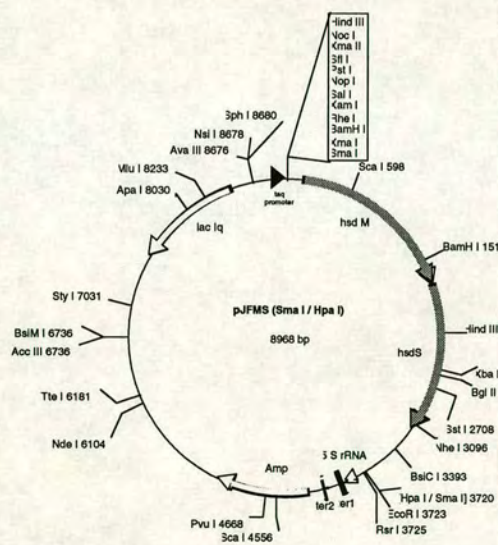
**a) 8026 bp *SmaI-EcoRI* DNA fragment inserted into pJF118HE**



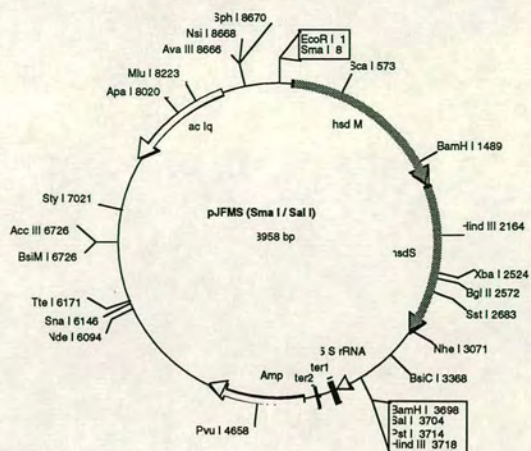
b) 3688 bp *SmaI-HpaI* DNA fragment inserted into pJF118HE

right way round

wrong way round



c) 3700 bp *Sma*I-*Sal*II DNA fragment  
inserted into pJF118EH





### 3.2.2. Construction of N- and C-terminal deletion derivatives

The m\* region is defined by the cluster of mutations associated with the *de novo* methylation phenotype, and therefore spans at least the amino acids Glu48 to Arg153 in HsdM. An *ScaI* site located near the 3'-end of the m\* region (corresponding to Tyr149 in HsdM) made it very easy to construct a precise C-terminal deletion derivative. Since the insertion of an *ScaI* fragment spanning the m\* region into the pJF118 expression system leads to products with slightly different C-terminal sequences depending on how the polylinker is oriented, both possible constructs were made (Fig. 3.2). The process also took advantage of an *ScaI* site located within the Ampicillin resistance gene.

The HsdM overexpressing plasmid pJF-M<sub>SmaI/HindIII</sub> (see Fig. 3.2.a; construct kindly provided by P. Thorpe) was cut with *ScaI*, and the 4985 bp DNA fragment containing the C-terminal half of the ampicillin resistance gene, the tac promoter, and the m\* region purified. pJF118HE was double digested with *SmaI* and *ScaI*, and the 836 bp fragment containing transcriptional terminators and the complementing N-terminal half of the ampicillin gene purified as well (Fig. 3.2.b). Ligation of the two fragments yields several products of which only the desired 5821 bp construct pJF-HE-m\* has a complete ampicillin resistance gene and could therefore easily be selected for. The same procedure with the equivalent 851 bp *SmaI/ScaI* fragment from pJF118EH resulted in an 5836 bp construct designated pJF-EH-m\* (Fig. 3.2.c).

The construction of N-terminal deletion derivatives requires an expression system that provides not only an inducible promoter, but an efficient translation initiation site and ATG start codon as well, and the well characterised T7 system was chosen for the task.

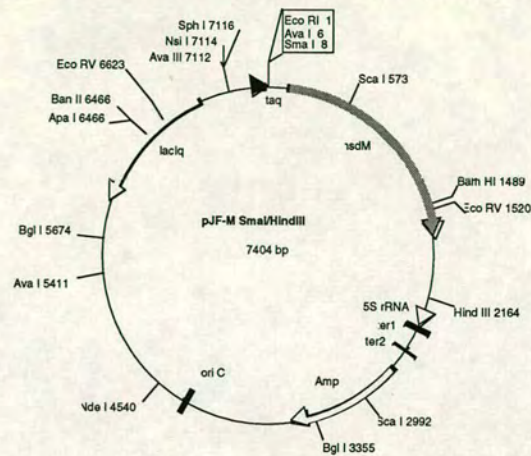
A 3123 bp *ScaI-HpaI* fragment containing the *hsdM* and *hsdS* genes without the m\* region was excised from pJFMS<sub>SmaI/EcoRI</sub> and inserted into the *SmaI* site of the 2488 bp plasmid pT7-7 (Fig. 3.3.a). The resulting 5611 bp construct pT7-7-MS( $\Delta$ m\*) (Fig. 3.3.b) encodes a truncated HsdM subunit with some additional amino acids encoded by the polylinker region at the N-terminus (Fig. 3.3.c).



**Fig. 3.2 Construction of C-terminal deletion derivatives**

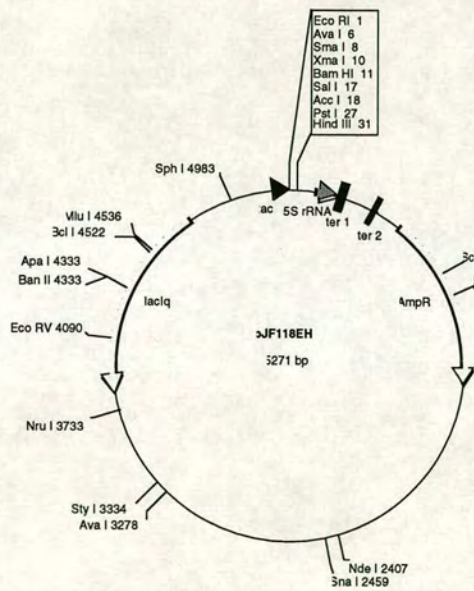
**a)**

**pJF-M<sub>Sma/HindIII</sub>**  
source of the 4985 bp *ScaI* fragment containing the m\* region

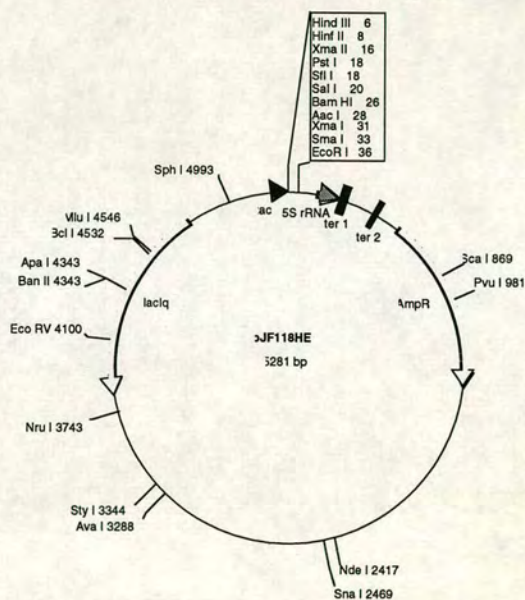


**b)**

**pJF118EH**  
source of the 851 bp *SmaI/ScaI* fragment

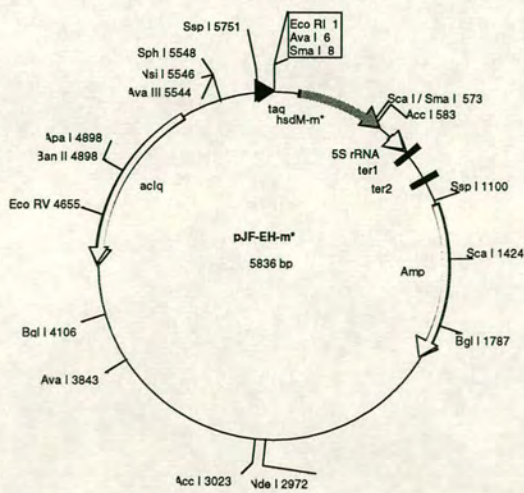


**pJF118HE**  
source of the 836 bp *SmaI/ScaI* fragment



**c)**

**pJF-EH-m\***  
5836 bp



**pJF-HE-m\***  
5821 bp

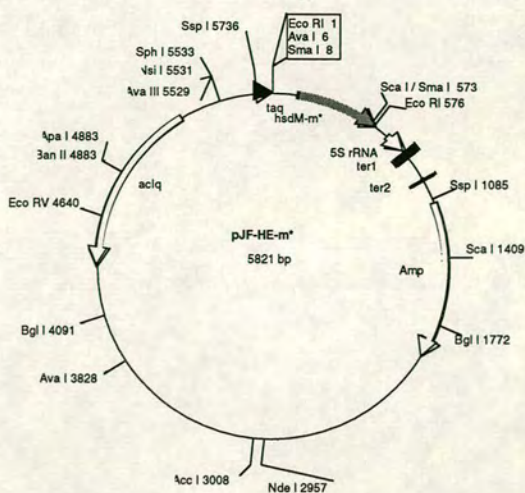
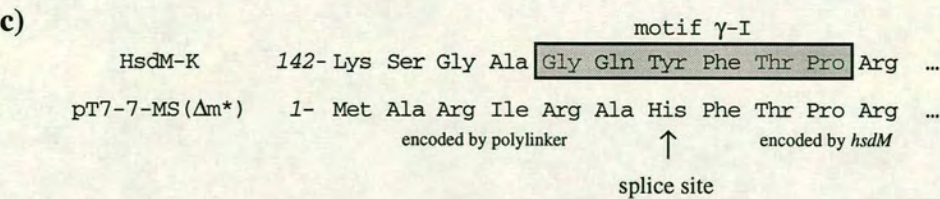
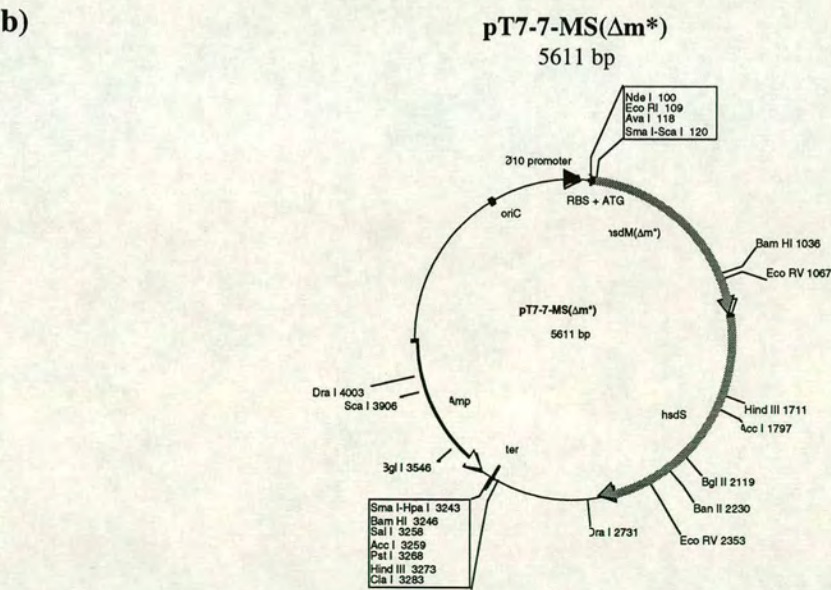
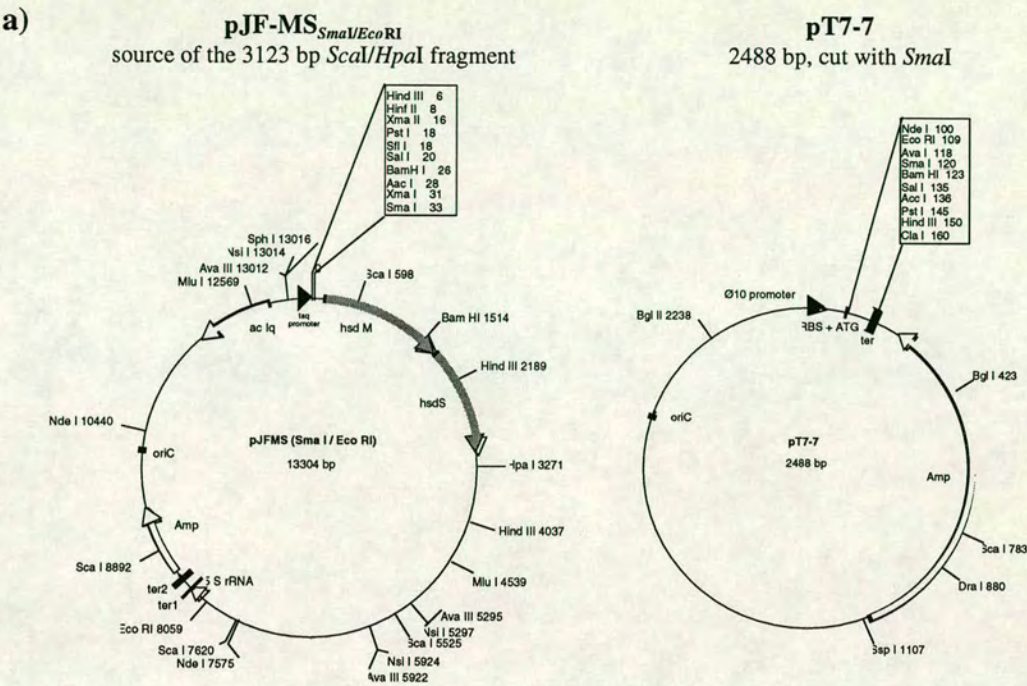




Fig. 3.3 Construction of N-terminal deletion derivatives





### 3.2.3. Identification of positive clones

The cloning and expression of constructs based on the pJF118 expression system was done in the bacterial strain NM679. The cloning of constructs based on the T7-7 system was also done in NM679, while expression of the T7-7 system was done in strains harbouring an IPTG inducible copy of the T7 polymerase gene (*e.g.* BL21 DE3, 784, 785a, 785b).

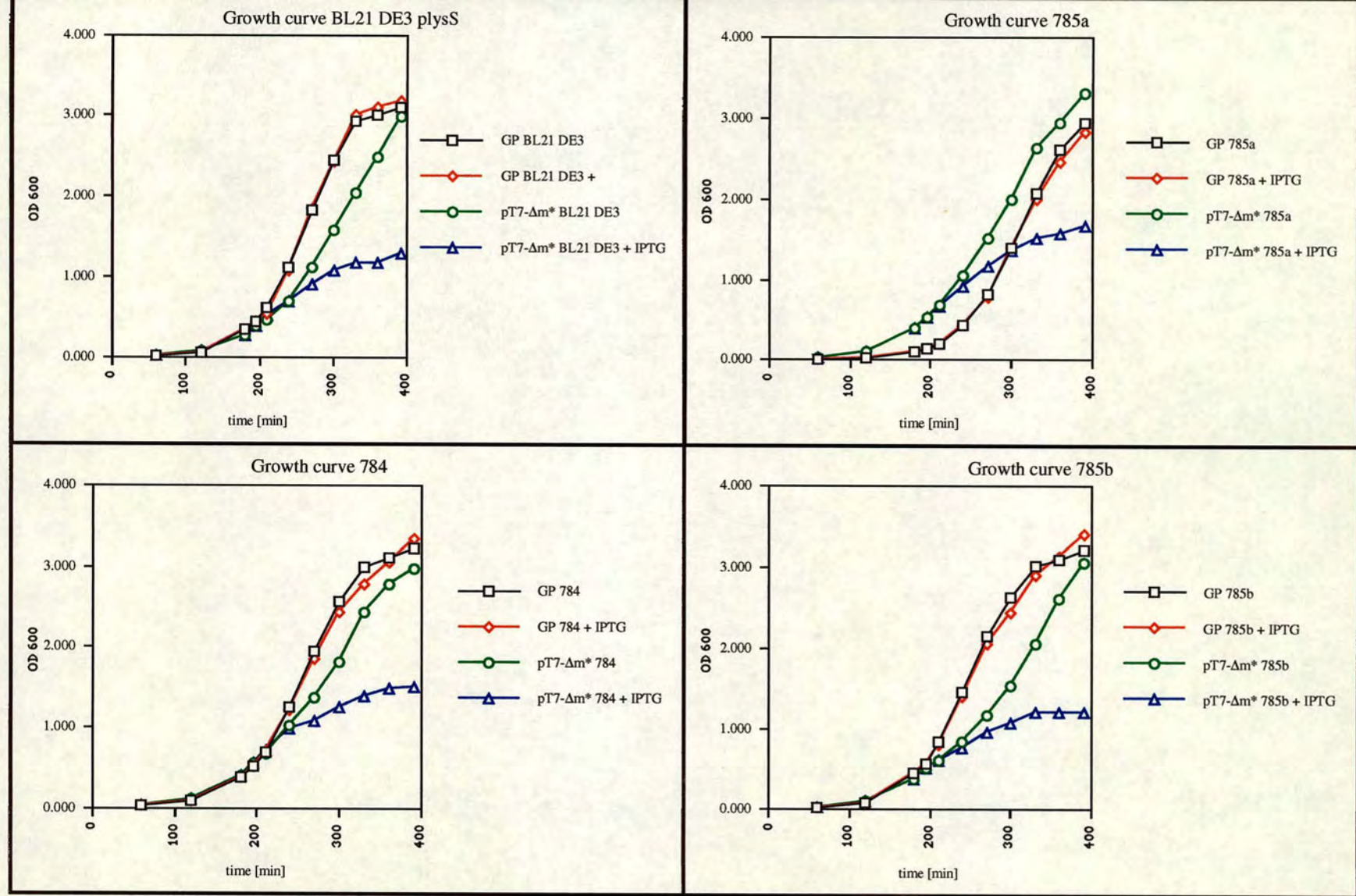
A specific probe for detecting the desired constructs was made by cutting pJF-M<sub>SmaI/HindIII</sub> (construct kindly provided by P. Thorpe) with *Bam*HI, purifying the 1488 bp fragment spanning the first half of the *hsdM* gene, and labeling it with digoxigenin. Positive transformants were identified by hybridisation with this *hsdM*-specific probe, and plasmid DNA prepared from each clone digested with *Bam*HI to determine the orientation of the insert and verify the construct.

### 3.3. Expression from the constructs

Regardless of whether the complete *hsdMS* genes were inserted into the pJF118 expression system as *SmaI-EcoRI*, *SmaI-HpaI*, or *SmaI-SalI* fragments, expression was strong enough to be detected on SDS-polyacrylamide gels by coomassie staining, and two of the clones (one r<sup>-</sup>, the other r<sup>+</sup>) were selected for purification. As with the wildtype methyltransferase the expressed proteins were found in the supernatant and therefore in soluble form.

The growth of cells induced for expression of plasmid pT7-7-MS( $\Delta$ m\*) harbouring *hsdMS* genes deleted for the N-terminal third of HsdM was strongly reduced compared to the appropriate controls, and came to a virtual stand-still about two hours after induction (see Fig. 3.4). SDS-polyacrylamide gelelectrophoresis (Fig. 3.5) revealed the strong expression of a product mainly found in the insoluble fraction and migrating at the same position in the gel as HsdS (molecular weight 51 402 D). It seems unlikely that this could be the product of the truncated *hsdM* gene, which codes for a polypeptide of molecular weight 43 102 D, and more likely that the truncated *hsdM* gene codes for an unstable HsdM subunit that is degraded in the cell, leaving the HsdS subunit behind.





**Fig. 3.4: Growth curves of pT7-7-MS( $\Delta m^*$ )**

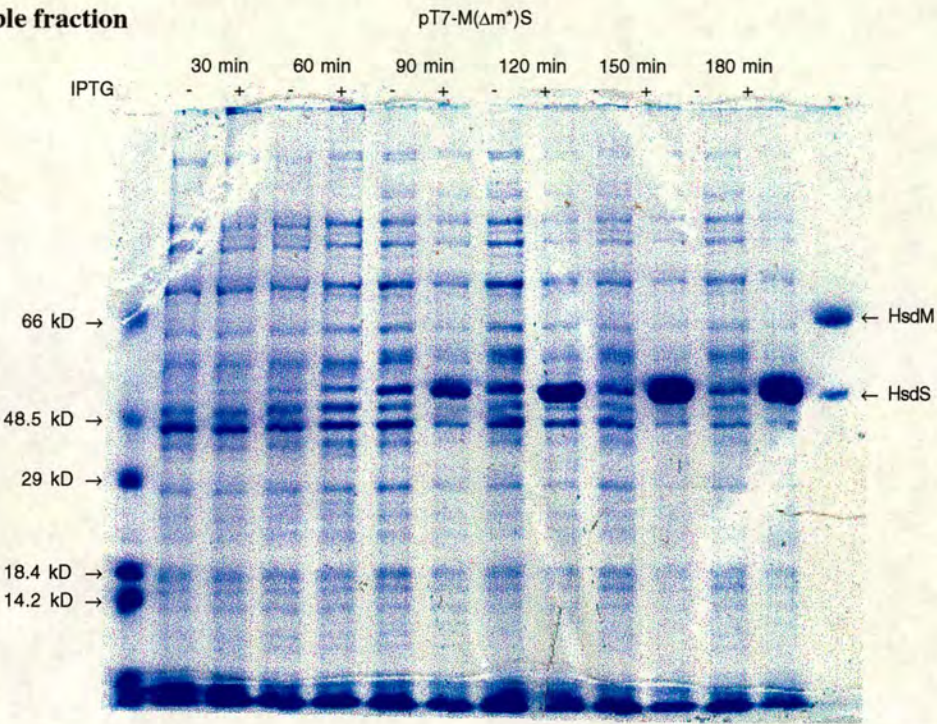
Strains harbouring pT7-7-MS( $\Delta m^*$ ) and induced for expression grew very poorly, and growth came to a virtual standstill about 2 hours after induction independent of the strain used. Controls with another pT7-7 derivative (construct kindly provided by Jill Smith) as well as uninduced strains harbouring the same plasmid grew normally.



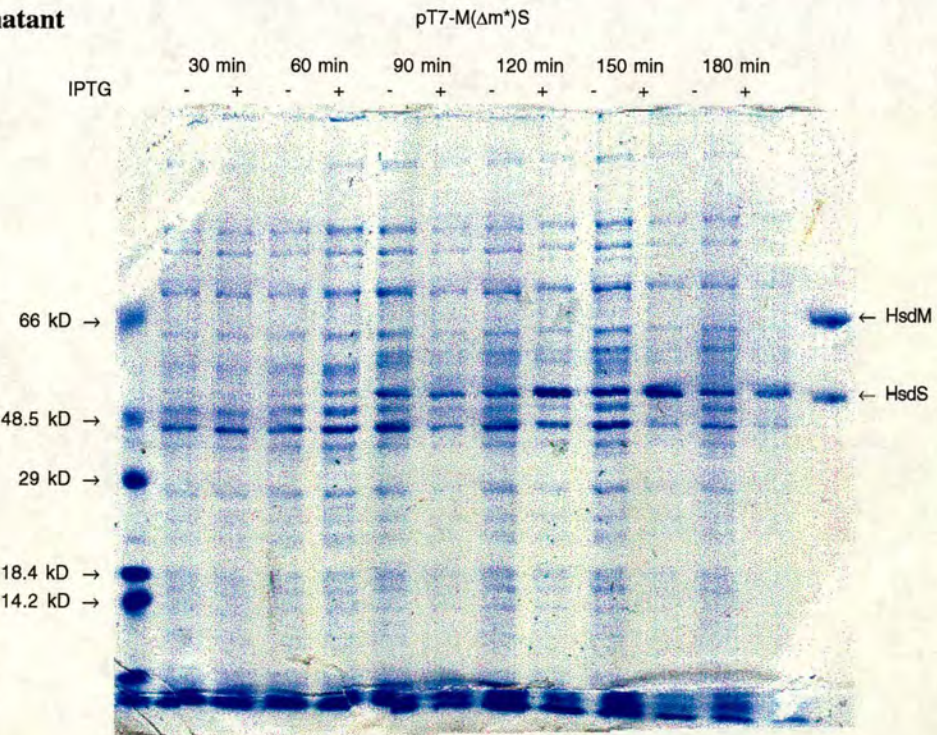
**Fig. 3.5: Expression of pT7-MS( $\Delta m^*$ )**

Induction of pT7-MS( $\Delta m^*$ ) with IPTG leads to the accumulation of a product of approx. 50 kD in the insoluble fraction.

**a) Insoluble fraction**



**b) Supernatant**





No expression could be detected on Coomassie stained SDS-polyacrylamide gels from either pJF-HE-m\* or pJF-EH-m\*, which encode the N-terminal third of HsdM on its own.

### 3.4. Purification

Two of the mutant methyltransferases were purified for further investigation (L134V which shows an r<sup>+</sup>m\* phenotype, and L113Q which exhibits an r<sup>-</sup>m\* phenotype; see table 2.2). In both cases the mutant methylases followed the elution profile seen for the wildtype methyltransferase and eluted at characteristic NaCl concentrations from the CM, Heparin, and DEAE columns used (fig. 3.6 and 3.7). Both mutant enzymes therefore appear to be very similar to the wildtype enzyme in their manner of purification, indicative of proteins which have no major structural differences.

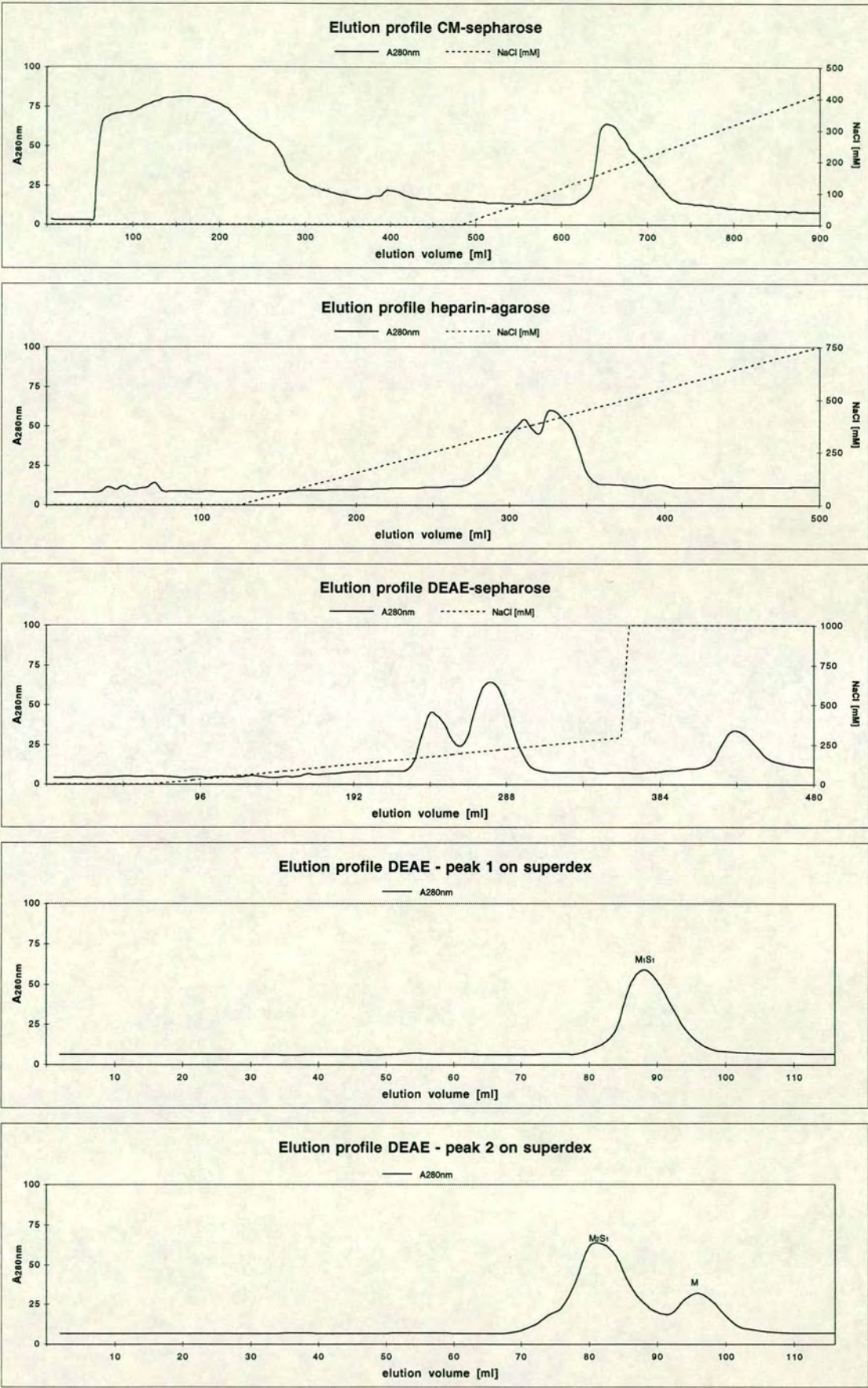
Protein eluted from the first column, CM-sepharose, as a broad peak at a salt concentration of between 0.1 and 0.2 M NaCl. Gradient elution from the second column, heparin-agarose, resulted in two protein peaks eluting at NaCl concentrations of approximately 0.23 - 0.27 M and 0.27 - 0.35 M. When the two peaks were pooled and applied to the third column, DEAE-sepharose, gradient elution again resulted in two protein peaks eluting at NaCl concentrations of approximately 0.23 - 0.3 M and 0.3 - 0.35 M respectively. SDS PAGE and gel filtration indicated the first of the two peaks to be predominantly of the form M<sub>1</sub>S<sub>1</sub>, with the second being predominantly M<sub>2</sub>S<sub>1</sub>, and the fractions from these two peaks were pooled separately.

When the second peak from the DEAE-sepharose column was applied onto a Hi-Load 16/60 Superdex 200 gel filtration column it eluted as two peaks. The first one of these eluted between 75 and 88 ml elution volume with a maximum peak at 81 ml, and was followed by a smaller second one eluting between 92 and 100 ml with the maximum at 96 ml elution volume. SDS PAGE showed the second of the two peaks to be HsdM, and indicated the first one to be of the form M<sub>2</sub>S<sub>1</sub>. Consistent with this the first peak from the DEAE-Sepharose applied onto the Superdex 200 gel filtration column eluted at an intermediate position from 82 - 96 ml with a maximum peak at 87 ml. All peaks eluting from the gel filtration column were pooled separately, their concentration determined, and kept in storage buffer at -20 °C until used.



**Fig. 3.6: Typical elution profiles in the purification of the EcoKI methyltransferase**

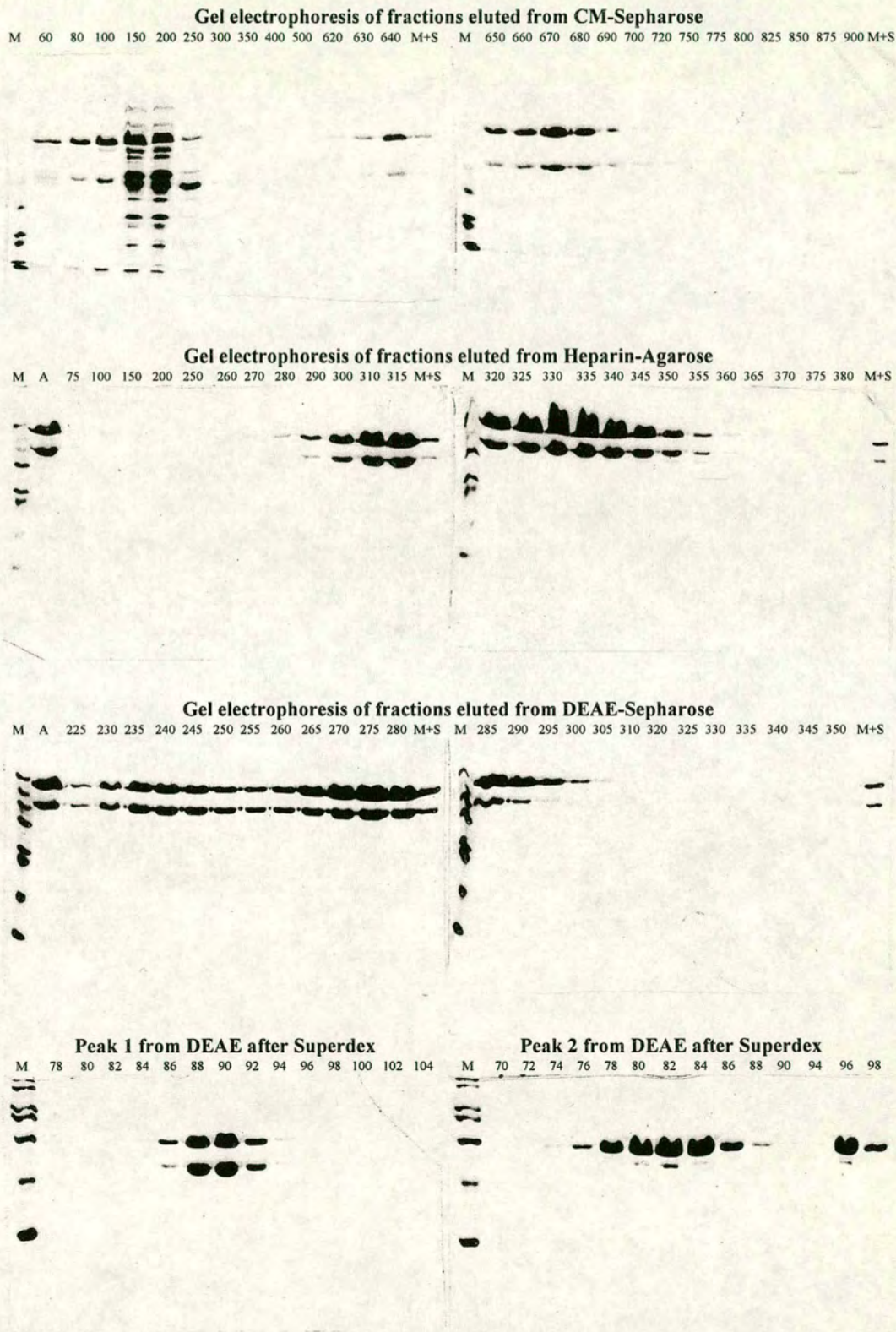
The elution profiles recorded for both the wildtype and the mutant methyltransferases by following the absorption at 280 nm were virtually identical.





**Fig. 3.7: Purification of the mutant L113Q**

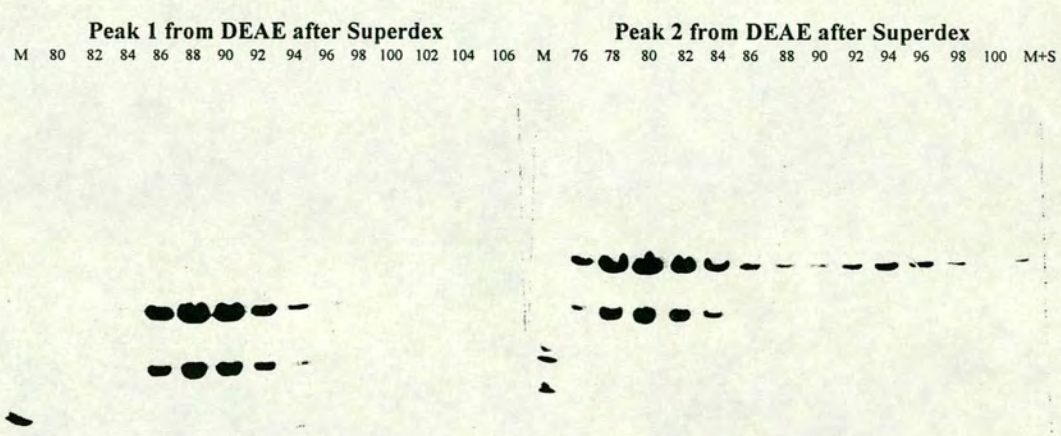
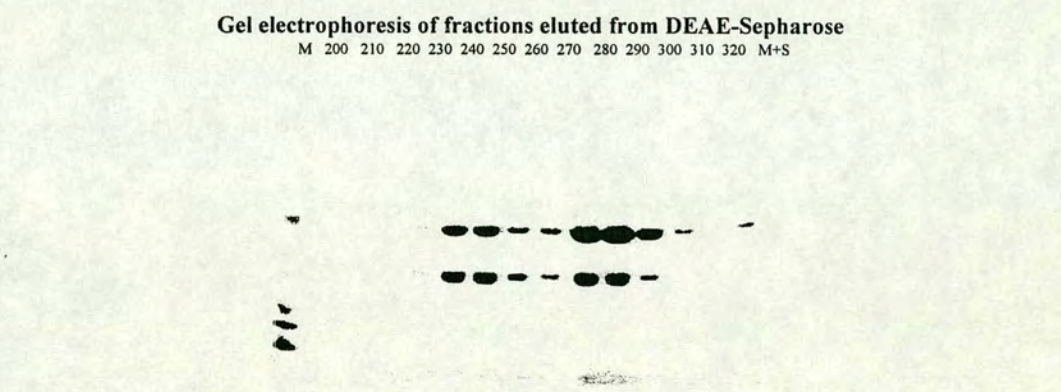
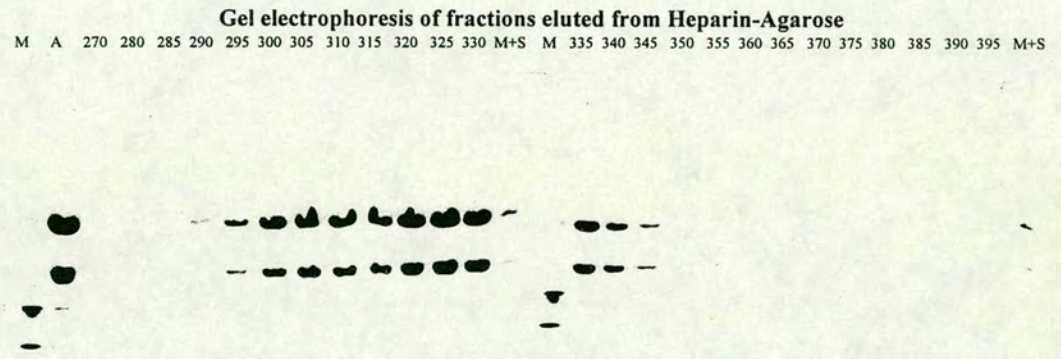
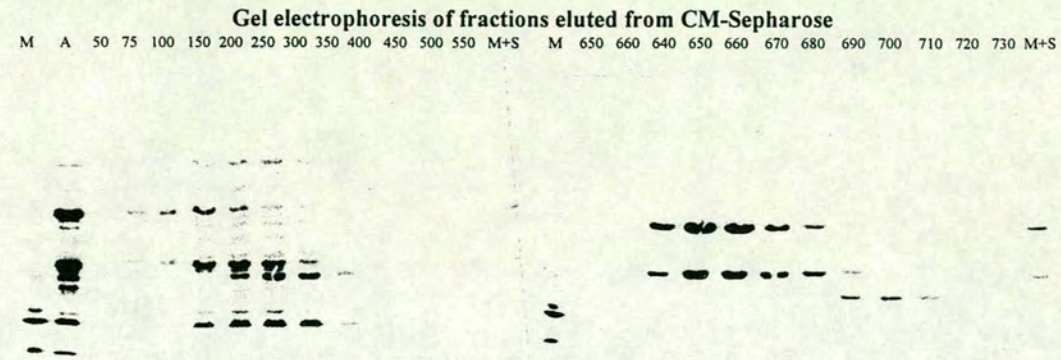
numbers represent the elution volume of the samples in ml A: Applied onto column M: Marker proteins M+S: control sample of M<sub>2</sub>S<sub>1</sub>





**Fig. 3.8: Purification of the mutant L134V**

numbers represent the elution volume of the samples in ml A: Applied onto column M: Marker proteins M+S: control sample of M<sub>2</sub>S<sub>1</sub>





# **CHAPTER 4:**

## **BIOCHEMICAL CHARACTERISATION**



#### 4.1. AdoMet Binding by Fluorescence of an extrinsic Fluorophore

Fluorescence is the phenomenon whereby a molecule, after absorbing radiation, emits radiation of a longer wavelength. At room temperature most organic molecules are in the ground state. Absorption of photons elevates an electron in these molecules to a higher energy state. After absorption, energy is lost very rapidly by collision degradation (as heat), resulting in the energy of the excited molecules falling rapidly to that of minimal vibrational energy in the lowest excited state. The energy emitted from these molecules on returning to the ground state gives rise to a fluorescent peak.

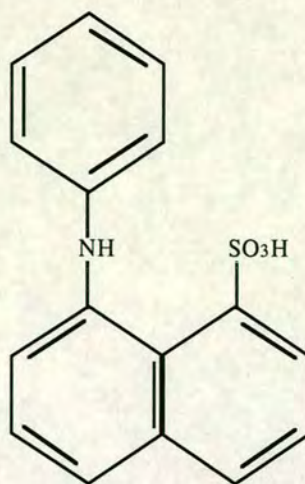
Although many organic molecules absorb in the ultraviolet and visible regions, only a few fluoresce, *i.e.* proteins that contain fluorescent chromophores such as tryptophan and FAD. Since the fluorescent properties of a molecule vary with its mobility and the polarity of its environment, such properties may be monitored in the vicinity of a fluorescent probe by measuring changes in its fluorescence. The binding of substances such as inhibitors, coenzymes, and allosteric effectors close to such intrinsic fluorophores can be measured by the change in the fluorescence spectrum they effect. This in turn may give information about the conformation, denaturation and aggregation of the proteins.

Studies like these may be extended to proteins lacking suitable intrinsic fluorophores by coupling them to extrinsic fluorophores such as ANS (see fig. 4.1) and measuring the so-called extrinsic fluorescence of the conjugate or complex. ANS contains both charged and hydrophobic areas and therefore locates at the water-lipid interface (*e.g.* membranes, nucleotide-binding sites and hydrophobic patches in proteins; Slavik, 1982; Chou *et al.*, 1989; Taylor *et al.*, 1992). The fluorescence of ANS in solution is very weak with an emission maximum at 515 nm, but is enhanced and shifted to shorter wavelengths upon binding to proteins.

The fluorescence of ANS is greater when bound to the methyltransferase than when in solution (Powell *et al.*, 1993), and its displacement upon addition of AdoMet to a solution containing both the methyltransferase and ANS can be followed by measuring the resulting decrease in fluorescence. This allows the determination of dissociation constants for AdoMet and the respective protein complex.



*Structure of ANS*



The data were gathered by titrating a solution containing both methyltransferase and ANS with increasing amounts of AdoMet, and the displacement of protein bound ANS by AdoMet monitored by fluorescence spectroscopy. AdoMet is fairly stable at low pH, but decomposes at pH 8 with a half-life of about 30 minutes (D.T.F. Dryden, personal communication). However the use of increasing concentrations of AdoMet made the problems associated with the instability of AdoMet at pH 8 irrelevant since the amount of decomposed AdoMet was insignificant compared to the amounts added at subsequent stages.

Dissociation constants were determined for the functional trimeric form of the wild-type and the two mutant methyltransferases L113Q and L134V as well as the  $M_1S_1$  complex of the wild-type and L113Q mutant by fitting different equations to the fluorescence data using the Grafit data analysis program.

The simplest one of these is the equation for ligand binding to a single site / two identical sites, and is of the form

$$y = \frac{[L] * Cap}{K_d + [L]} + background$$

where y is either the amount bound or is some factor proportional to it, *e.g.* the decrease in fluorescence. The capacity for binding ligand, Cap, is the amount bound in terms of moles of ligand.  $K_d$  is the dissociation constant of the ligand, and [L] is the concentration of free ligand.

Both HsdM subunits are the same in the EcoKI methyltransferase, and data for the wild-type enzyme had indicated that the two binding sites for AdoMet bind the



cofactor in an identical manner. However it could not be assumed that the mutants would behave in the same way - although it was difficult to imagine how a mutation might affect AdoMet binding in one subunit but not the other. Therefore the equation for two non-identical sites

$$y = \frac{[L] * Cap_1}{K_{d1} + [L]} + \frac{[L] * Cap_2}{K_{d2} + [L]} + \text{background}$$

in which the two sites are characterised by individual  $K_d$  and Cap values, was also applied.

Since methylation at one half target site of the recognition sequence is dependent on the methylation state of the other half target site, both HsdM subunits have to be able to communicate with each other. It is therefore possible that binding of AdoMet to one HsdM subunit affects binding of AdoMet by the other subunit, and the equation for co-operative binding of ligand to two identical sites

$$y = \frac{[L]^n * Cap_1}{(K_{d1} + [L])^n} + \text{background}$$

where  $n$  is a measure of the cooperativity, was fitted to the data too. For normal binding to a single site,  $n$  should not differ significantly from 1.0.

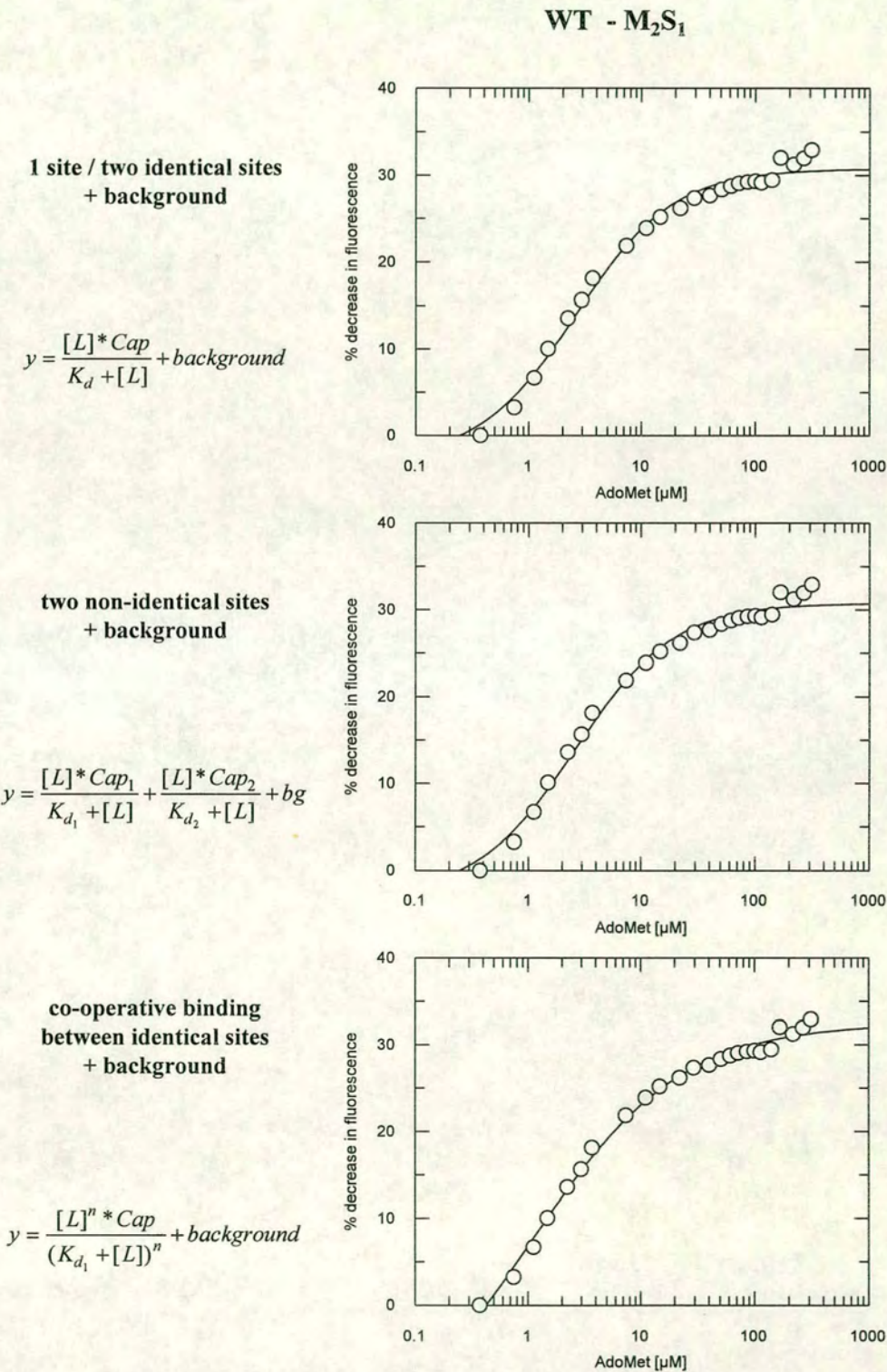
One particular and unexplained problem with the assay was that the fluorescence of the methyltransferase / ANS solution would steadily decrease with the addition of AdoMet except for the very first addition of the cofactor. The first addition of AdoMet to the solution could instead of a decrease even lead to an increase in the fluorescence emission. Since the fluorescence emission would thereafter continually decline, the first value of the data set (corresponding to zero concentration of AdoMet) was omitted from subsequent calculations. Instead the next value at 0.375  $\mu$ M AdoMet was used as reference for the decrease in the fluorescence emission. To account for this versions of the respective equations incorporating a background value were used.

Figures 4.2. a) - e) demonstrate the decrease in fluorescence observed when a solution containing 1  $\mu$ M methyltransferase and 50  $\mu$ M ANS is titrated with AdoMet. Tables 4.2. a) - e) detail the dissociation constants ( $K_d$  values) for AdoMet obtained from these curves and the significance analysis of the equations fitted.



**Fig. 4.2 a) AdoMet Binding by Methyltransferase as determined by Titration of Methyltransferase + ANS with AdoMet.**

The displacement of bound ANS and consequent reduction in fluorescence as a function of AdoMet concentration.





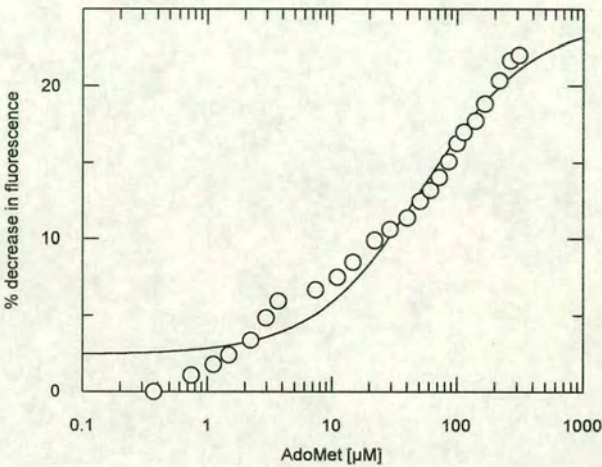
**Fig. 4.2 b) AdoMet Binding by Methyltransferase as determined by Titration of Methyltransferase + ANS with AdoMet.**

The displacement of bound ANS and consequent reduction in fluorescence as a function of AdoMet concentration.

**L113Q - M<sub>2</sub>S<sub>1</sub>**

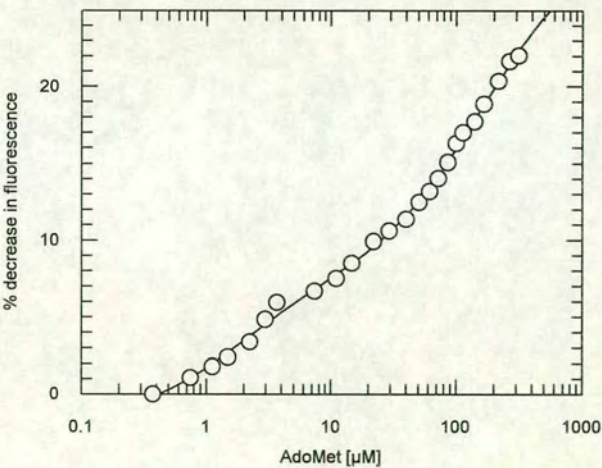
**1 site / two identical sites  
+ background**

$$y = \frac{[L] * Cap}{K_d + [L]} + background$$



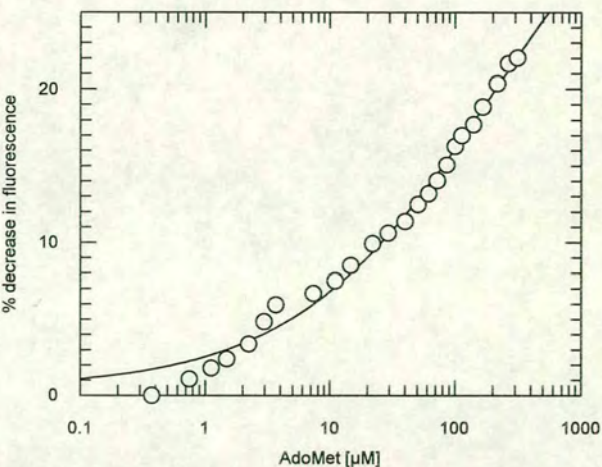
**two non-identical sites  
+ background**

$$y = \frac{[L] * Cap_1}{K_{d1} + [L]} + \frac{[L] * Cap_2}{K_{d2} + [L]} + bg$$



**co-operative binding  
between identical sites  
+ background**

$$y = \frac{[L]^n * Cap}{(K_{d1} + [L])^n} + background$$





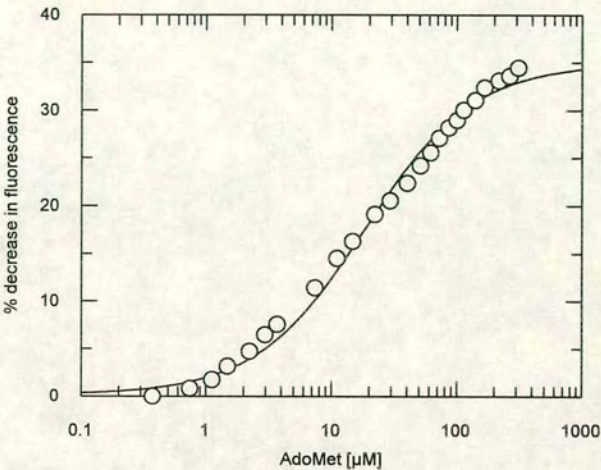
**Fig. 4.2 c) AdoMet Binding by Methyltransferase as determined by Titration of Methyltransferase + ANS with AdoMet.**

The displacement of bound ANS and consequent reduction in fluorescence as a function of AdoMet concentration.

**L134V - M<sub>2</sub>S<sub>1</sub>**

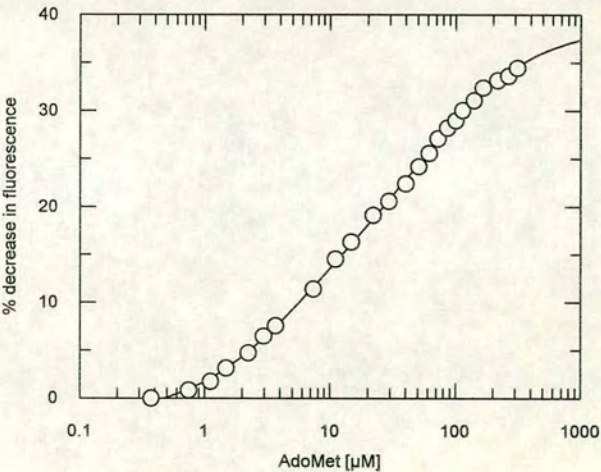
**1 site / two identical sites  
+ background**

$$y = \frac{[L] * Cap}{K_d + [L]} + background$$



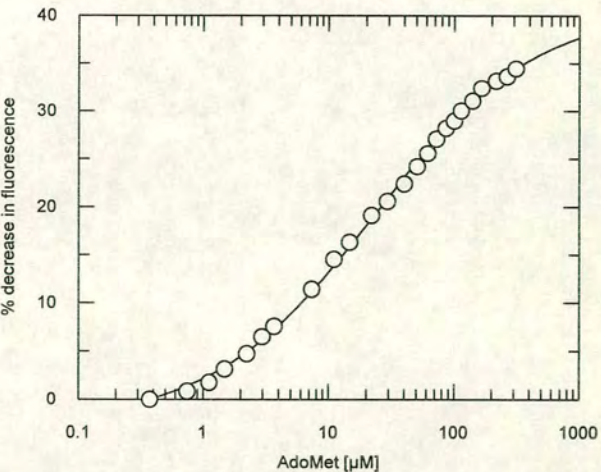
**two non-identical sites  
+ background**

$$y = \frac{[L] * Cap_1}{K_{d_1} + [L]} + \frac{[L] * Cap_2}{K_{d_2} + [L]} + bg$$



**co-operative binding  
between identical sites  
+ background**

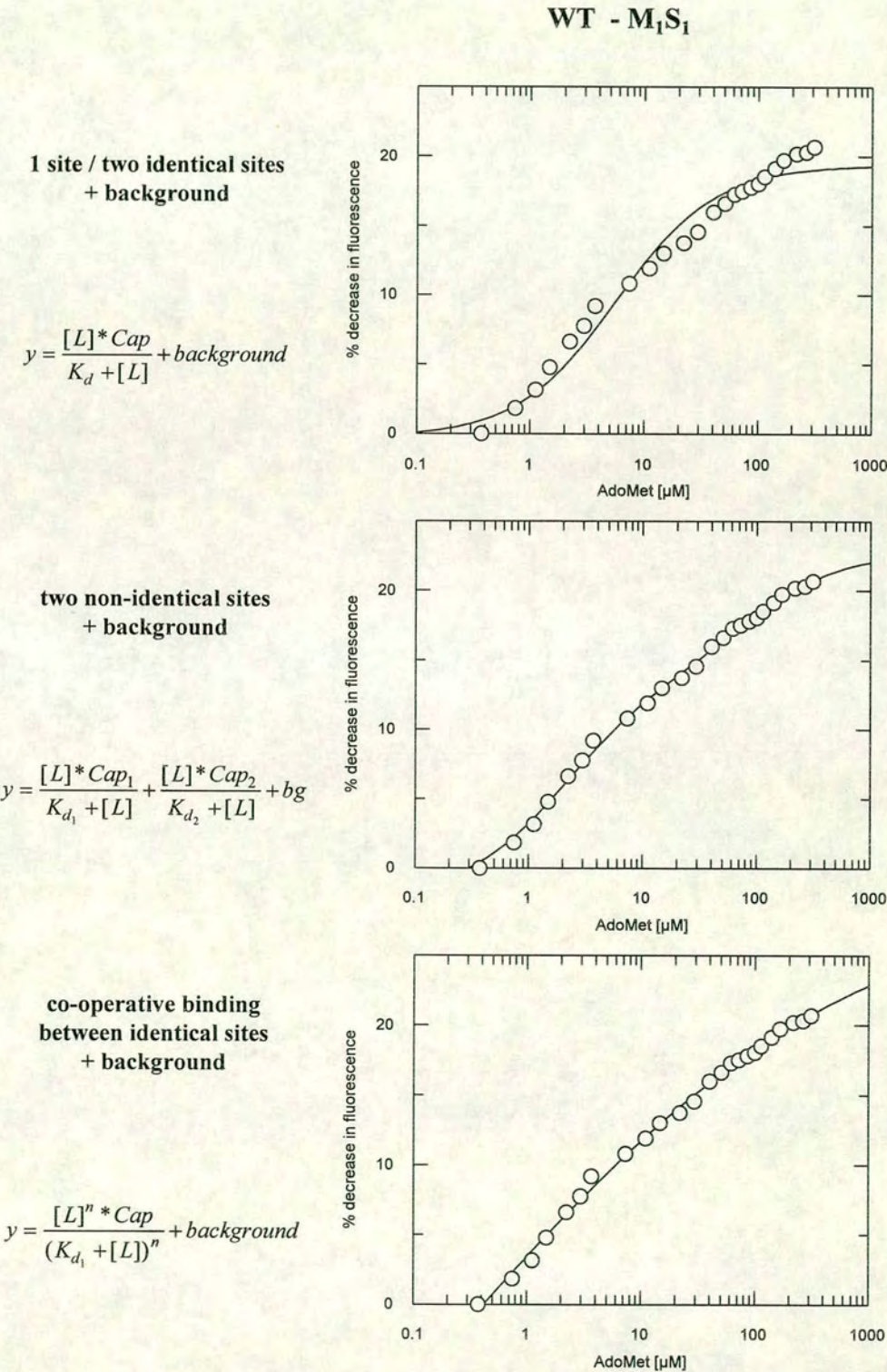
$$y = \frac{[L]^n * Cap}{(K_{d_1} + [L])^n} + background$$





**Fig. 4.2 d) AdoMet Binding by Methyltransferase as determined by Titration of Methyltransferase + ANS with AdoMet.**

The displacement of bound ANS and consequent reduction in fluorescence as a function of AdoMet concentration.





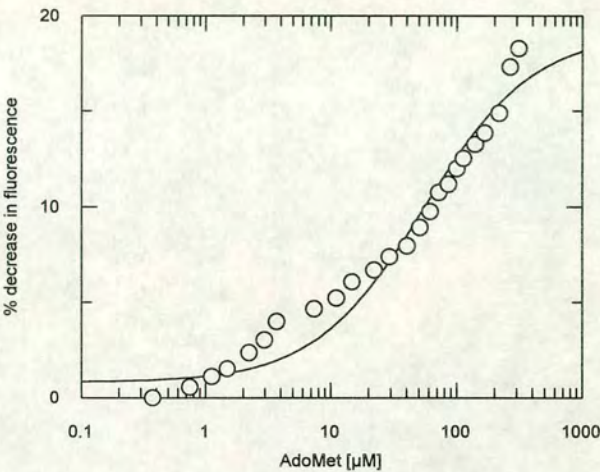
**Fig. 4.2 e) AdoMet Binding by Methyltransferase as determined by Titration of Methyltransferase + ANS with AdoMet.**

The displacement of bound ANS and consequent reduction in fluorescence as a function of AdoMet concentration.

**L113Q - M<sub>1</sub>S<sub>1</sub>**

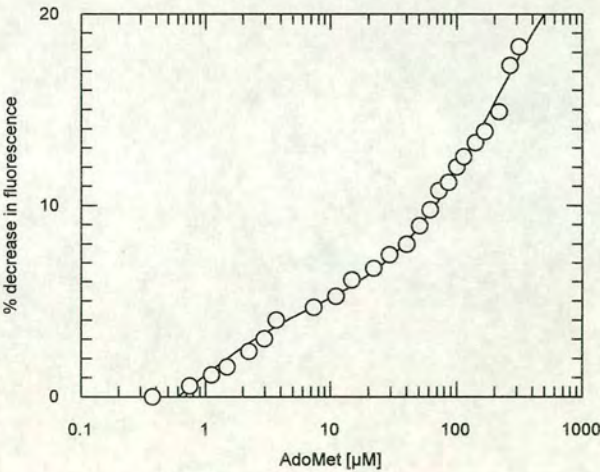
**1 site / two identical sites  
+ background**

$$y = \frac{[L] * Cap}{K_d + [L]} + background$$



**two non-identical sites  
+ background**

$$y = \frac{[L] * Cap_1}{K_{d_1} + [L]} + \frac{[L] * Cap_2}{K_{d_2} + [L]} + bg$$



**co-operative binding  
between identical sites  
+ background**

$$y = \frac{[L]^n * Cap}{(K_{d_1} + [L])^n} + background$$

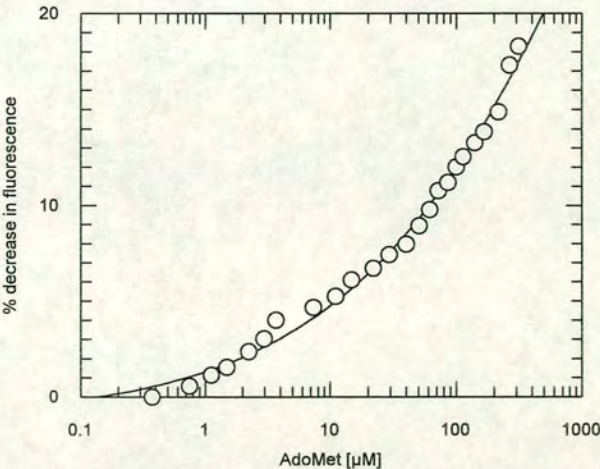




Table. 4.2 a) AdoMet Binding by Methyltransferase as determined by Titration of Methyltransferase + ANS with AdoMet.

WT - M <sub>2</sub> S <sub>1</sub>		value	standard error
1 site / two identical sites + background $y = \frac{[L]*Cap}{K_d + [L]} + background$	reduced $\chi^2$ :	1.229	
	background :	-2.89	0.86
	Capacity :	33.66	0.85
	K <sub>d</sub> :	2.62	0.24
two non-identical sites + background $y = \frac{[L]*Cap_1}{K_{d1} + [L]} + \frac{[L]*Cap_2}{K_{d2} + [L]} + bg$	reduced $\chi^2$ :	1.351	
	background :	-2.92	0.84
	Capacity <sub>1</sub> :	16.85	2.10
	K <sub>d1</sub> :	2.23	1.62
	Capacity <sub>2</sub> :	16.85	2.10
	K <sub>d2</sub> :	3.06	1.98
co-operative binding between identical sites + background $y = \frac{[L]^n * Cap}{(K_{d1} + [L])^n} + background$	reduced $\chi^2$ :	0.7106	
	background :	-17.82	7.12
	Capacity :	50.32	7.70
	K <sub>d</sub> :	1.04	0.30
	cooperativity :	0.65	0.09



Table. 4.2 b) AdoMet Binding by Methyltransferase as determined by Titration of Methyltransferase + ANS with AdoMet.

L113Q - M <sub>2</sub> S <sub>1</sub>		value	standard error
1 site / two identical sites + background $y = \frac{[L] * Cap}{K_d + [L]} + background$	reduced $\chi^2$ :	1.128	
	background :	2.44	0.43
	Capacity :	21.83	1.02
	K <sub>d</sub> :	53.37	8.61
two non-identical sites + background $y = \frac{[L] * Cap_1}{K_{d1} + [L]} + \frac{[L] * Cap_2}{K_{d2} + [L]} + bg$	reduced $\chi^2$ :	0.08218	
	background :	-1.87	0.54
	Capacity <sub>1</sub> :	22.68	0.83
	K <sub>d1</sub> :	172.19	22.14
	Capacity <sub>2</sub> :	9.61	0.47
	K <sub>d2</sub> :	1.79	0.37
co-operative binding between identical sites + background $y = \frac{[L]^n * Cap}{(K_{d1} + [L])^n} + background$	reduced $\chi^2$ :	0.6466	
	background :	0.48	0.68
	Capacity :	39.31	10.23
	K <sub>d</sub> :	17.97	3.21
	cooperativity :	0.54	0.09



Table. 4.2 c) AdoMet Binding by Methyltransferase as determined by Titration of Methyltransferase + ANS with AdoMet.

L134V - M <sub>2</sub> S <sub>1</sub>		value	standard error
1 site / two identical sites + background $y = \frac{[L]*Cap}{K_d + [L]} + background$	reduced $\chi^2$ :	1.312	
	background :	0.18	0.53
	Capacity :	34.77	0.68
	K <sub>d</sub> :	18.41	1.70
two non-identical sites + background $y = \frac{[L]*Cap_1}{K_{d_1} + [L]} + \frac{[L]*Cap_2}{K_{d_2} + [L]} + bg$	reduced $\chi^2$ :	0.08383	
	background :	-1.87	0.22
	Capacity <sub>1</sub> :	18.40	1.33
	K <sub>d1</sub> :	4.58	0.57
	Capacity <sub>2</sub> :	22.40	1.10
	K <sub>d2</sub> :	69.18	9.58
co-operative binding between identical sites + background $y = \frac{[L]^n * Cap}{(K_{d_1} + [L])^n} + background$	reduced $\chi^2$ :	0.1615	
	background :	-2.49	0.37
	Capacity :	43.38	1.09
	K <sub>d</sub> :	8.35	0.51
	cooperativity :	0.67	0.02



Table. 4.2 d) AdoMet Binding by Methyltransferase as determined by Titration of Methyltransferase + ANS with AdoMet.

WT - M <sub>1</sub> S <sub>1</sub>		value	standard error
1 site / two identical sites + background $y = \frac{[L]*Cap}{K_d + [L]} + background$	reduced $\chi^2$ :	1.197	
	background :	-0.26	0.69
	Capacity :	19.66	0.69
	K <sub>d</sub> :	5.67	0.88
two non-identical sites + background $y = \frac{[L]*Cap_1}{K_{d1} + [L]} + \frac{[L]*Cap_2}{K_{d2} + [L]} + bg$	reduced $\chi^2$ :	0.1166	
	background :	-2.00	0.30
	Capacity <sub>1</sub> :	15.49	0.66
	K <sub>d1</sub> :	2.04	0.23
	Capacity <sub>2</sub> :	9.33	0.54
	K <sub>d2</sub> :	84.15	23.09
co-operative binding between identical sites + background $y = \frac{[L]^n * Cap}{(K_{d1} + [L])^n} + background$	reduced $\chi^2$ :	0.1234	
	background :	-33.12	14.08
	Capacity :	63.31	16.44
	K <sub>d</sub> :	0.73	0.22
	cooperativity :	0.25	0.05



Table. 4.2 e) AdoMet Binding by Methyltransferase as determined by Titration of Methyltransferase + ANS with AdoMet.

L113Q - M <sub>1</sub> S <sub>1</sub>		value	standard error
1 site / two identical sites + background $y = \frac{[L]*Cap}{K_d + [L]} + background$	reduced $\chi^2$ :	2.263	
	background :	0.83	0.57
	Capacity :	18.25	1.44
	K <sub>d</sub> :	54.95	14.80
two non-identical sites + background $y = \frac{[L]*Cap_1}{K_{d1} + [L]} + \frac{[L]*Cap_2}{K_{d2} + [L]} + bg$	reduced $\chi^2$ :	0.1947	
	background :	-3.87	0.423
	Capacity <sub>1</sub> :	21.50	1.57
	K <sub>d1</sub> :	205.16	37.75
	Capacity <sub>2</sub> :	8.66	0.52
	K <sub>d2</sub> :	0.77	0.15
co-operative binding between identical sites + background $y = \frac{[L]^n * Cap}{(K_{d1} + [L])^n} + background$	reduced $\chi^2$ :	0.2541	
	background :	-1.14	0.33
	Capacity :	66.26	5.28
	K <sub>d</sub> :	25.87	2.43
	cooperativity :	0.40	0.02



Three different equations were fitted to the data. The first one is applicable to proteins with either a single binding site or two identical sites, while the second one applies to proteins with two different binding sites or a mixture of two different proteins competing for binding to a single site (or different forms of the same protein, each with its own  $K_d$ ). The third equation fits to proteins that show co-operative behaviour in the binding of a ligand to two identical binding sites.

The  $M_2S_1$  complex of the wild-type methyltransferase had previously been shown by equilibrium dialysis to have two AdoMet-binding sites which are either identical or so similar that no co-operative interactions between them can be discovered by the method used (Powell *et al.*, 1993). The dissociation constant was determined to be  $3.60 (\pm 0.42) \mu M$  by equilibrium dialysis, or  $2.21 (\pm 0.29) \mu M$  by ANS titration.

The results from the ANS titration of the  $M_2S_1$  complex of the wild-type methyltransferase are in excellent agreement with the published results. The  $K_d$  is  $2.62 (\pm 0.24) \mu M$  when the equation for two identical sites is fitted to the curve, and a similar  $K_d$  is attained when the equation for two non-identical sites is fitted. Both sites are therefore indistinguishable in the wild-type.

The titration of the mutant  $M_2S_1$  complexes shows that the mutants have a lower affinity for AdoMet than the wild-type. Fitting the equation for two identical sites results in a rather poor fit compared to the fit obtained with the equation for two non-identical sites. This indicates the existence of two different binding sites on the protein, one with a  $K_d$  very similar to the one determined for the trimeric enzyme, the other one much higher. While this could mean the existence of additional AdoMet binding sites on the methyltransferase, the result can equally be explained by proposing two different possible conformations of the HsdM subunit, one with high affinity for AdoMet, the other with a low affinity. That the L113Q mutant binds AdoMet with an even lower affinity than the L134V mutant might mean that a higher proportion of the protein adopts the "low affinity conformation".

The titrations were repeated with the  $M_1S_1$  complex of the wild-type and the mutant L113Q methyltransferase (the  $M_1S_1$  complex of the L134V mutant was accidentally lost during preparation). Again the equation for ligand-binding to one site fits significantly worse to the data than the equation for two different binding sites. While this is consistent with the existence of two different forms of the protein, it also shows that the observed effect is an inherent property of the protein and does not depend on interactions between the two HsdM subunits.



Although the equation for "co-operative binding" results in a curve that fits the data quite well, the fits are not significantly better than the ones assuming two different binding sites. Furthermore the cooperativity index is not significantly different from 1.0, the value for normal binding to a single site. Since it is difficult to imagine how the binding curve seen for the  $M_1S_1$  complex of both the wild-type and the L113Q mutant methyltransferase can be explained by "co-operative binding", fitting the equation for two non-identical binding sites to the data seems appropriate.

The consequence of the high  $K_d$  and therefore low affinity for AdoMet in one HsdM subunit means that AdoMet is only bound at one of the two HsdM subunits in the mutant  $M_2S_1$  complexes. The implications of these findings for the mechanism of methylation by the *EcoKI* methyltransferase are discussed in chapter 6.



#### 4.2. *In vitro* Methylase Activity Assay (by fluorographic Detection of [methyl-<sup>3</sup>H] Transfer).

The Type IA and IC families are known to have a preference for hemi-methylated DNA in the methylation reaction (Suri *et al.*, 1984; Dryden *et al.*, 1993; Taylor *et al.*, 1993) However the investigation of enzyme activities and substrate preferences *in vitro* which became possible with the availability of large amounts of highly purified methyltransferase proved unexpectedly difficult.

To investigate the *in vitro* activity of the *Eco*KI methyltransferase, a modification of an assay originally developed by Rubin and Modrich (1977) and involving scintillation counting of tritiated DNA substrate immobilised on ion-exchange paper had been used by Dryden and Cooper (1993). The methyltransferase was assayed for more than an hour at 25 °C in a reaction mixture containing 0.5 µM methyltransferase, 3 µM [methyl-<sup>3</sup>H]AdoMet and 1.5 µM DNA oligonucleotide duplex. Since this assay did not seem sensitive enough to detect enzyme activity at low enzyme concentrations, alternative assays were developed (see for example Willcock *et al.*, 1994).

The major obstacles in the adaptation of Rubin and Modrich's assay seemed to be the high concentration of methylase and the relatively low concentrations of DNA and AdoMet used, the low activity displayed by the methyltransferase (as seen in the long incubation periods), and the instability and high cost of [methyl-<sup>3</sup>H]AdoMet. Two of the problems could immediately be overcome by spiking highly concentrated and cheap unlabelled AdoMet with [methyl-<sup>3</sup>H]AdoMet, making it possible to vary the AdoMet concentration in the assay from 1 to 150 µM. However no activity was seen at low enzyme concentrations of between 5 and 100 nM. Consequently the enzyme concentration was increased to above 100 nM, and the DNA concentration raised to up to 10 µM. The resulting increase in the rate of methylation made a quantitative and accurate investigation of the parameters in the methylation reaction possible.

Rates of enzymatic reactions are usually measured under steady state conditions in which the concentrations of intermediates, namely that of the enzyme-substrate complex, stay the same while the concentrations of free substrates and products change slowly. As a consequence there is a period of time when product formation occurs at a linear rate, which makes it easy to measure the reactions velocity and deduce the kinetic parameters  $K_M$  and  $v_{max}$  of the reaction. Steady state conditions



therefore usually require low enzyme concentrations and the substrates to be in vast (at least 10fold) excess over the enzyme, with the substrate levels being varied around their respective  $K_m$ . However although the enzyme concentration in the assays were high, the rate of product formation remained linear for some time (possibly due to the slow speed of the methylation reaction) before levelling off, allowing the determination of initial rates in the reaction.

The standard assay used 0.555  $\mu\text{M}$  methyltransferase, 50  $\mu\text{M}$  AdoMet, and 10  $\mu\text{M}$  25mer oligonucleotide duplex DNA - M3105/M3106 as unmodified DNA, M3105/N8625 as methylated in the lower strand, and N8624/M3106 as methylated in the upper strand (the oligonucleotides P8870 - P8873 were identical replacements; see table 2.4 for details). To determine the  $K_m$  and  $v_{\text{max}}$  values for AdoMet and DNA in the methylation reaction, the standard assay was used but the concentration of the substrate under investigation varied. Concentrations were varied from 1 to 150  $\mu\text{M}$  for AdoMet, and from 1 to 20  $\mu\text{M}$  for DNA.

To investigate why no enzyme activity was found at low enzyme concentrations, the influence of the methyltransferase concentration on methylation was also examined by varying the methylase concentration in the standard assay from 0.111 to 1.1  $\mu\text{M}$ .

Aliquots were removed from the reaction mixture at different times, the reaction stopped by addition of  $\text{NH}_4\text{HCO}_3$ , and the amount of methylation determined by scintillation counting. The counts per minute were converted into pmol AdoMet, plotted against the time, and the initial velocities for the methylation reaction determined by linear regression. The resulting values (pmol AdoMet used per minute) were plotted against the substrate concentration and fitted to the standard Michaelis-Menten equation:

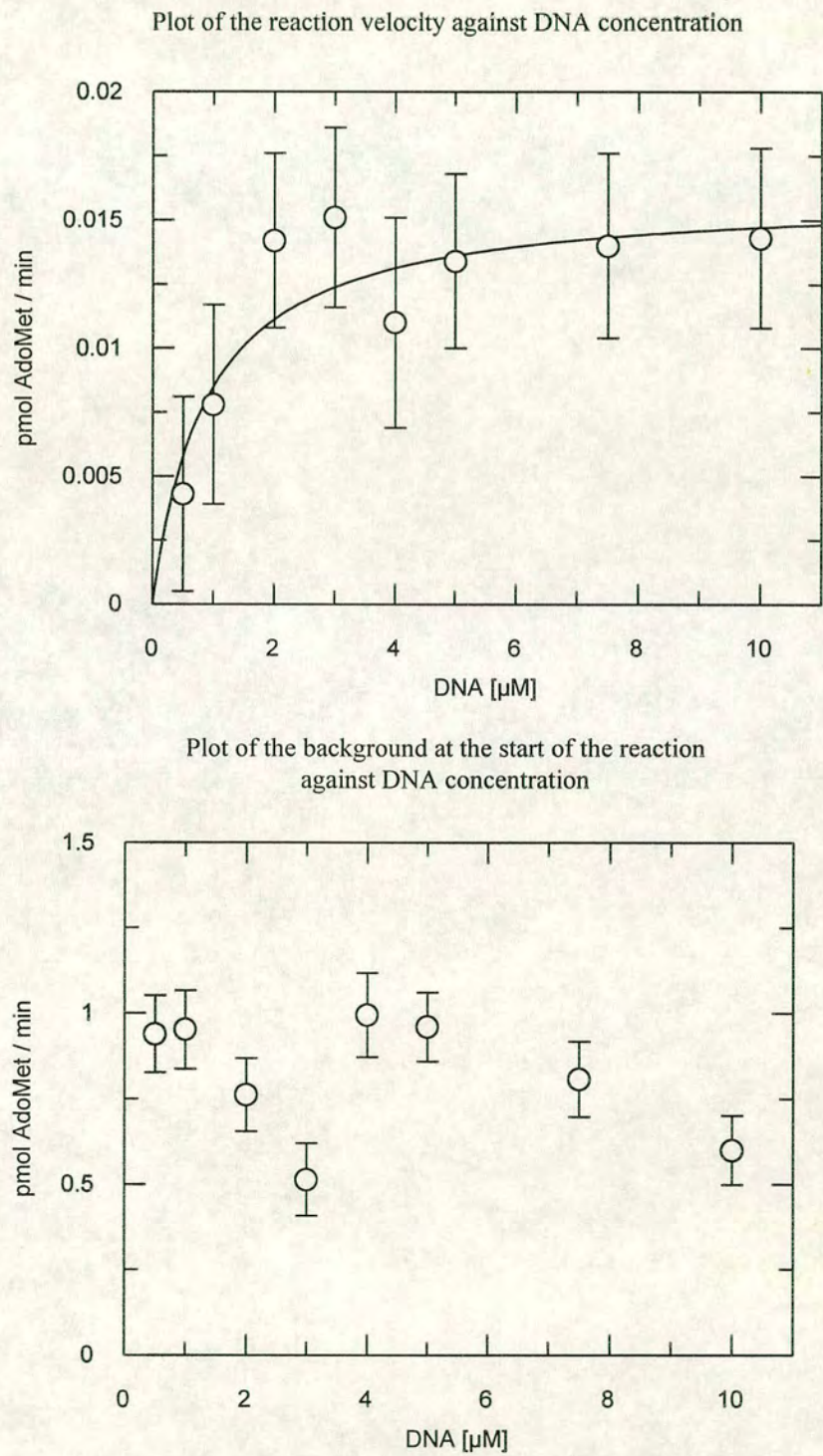
$$v = \frac{V_{\text{max}} * [\text{S}]}{K_M + [\text{S}]}$$

Under ideal conditions the linear regression analysis should result in a straight line through the point of origin, meaning of course that no transfer of methyl groups had occurred at time zero. However while the background level for scintillation counting had been accounted for (by measuring vials with a piece of paper in scintillation fluid) and subtracted from the data, some of the assays showed a considerable deviation from the expected result. To investigate this further the intercept with the y-axis (equivalent to pmol AdoMet "transferred" at time zero) was also plotted against the substrate concentration.



**Fig. 4.3.1. *In vitro* Assay of *M.Eco* KI methyltransferase activity by fluorographic Detection of [methyl-<sup>3</sup>H] Transfer - Dependence of the methylation reaction on DNA concentration.**

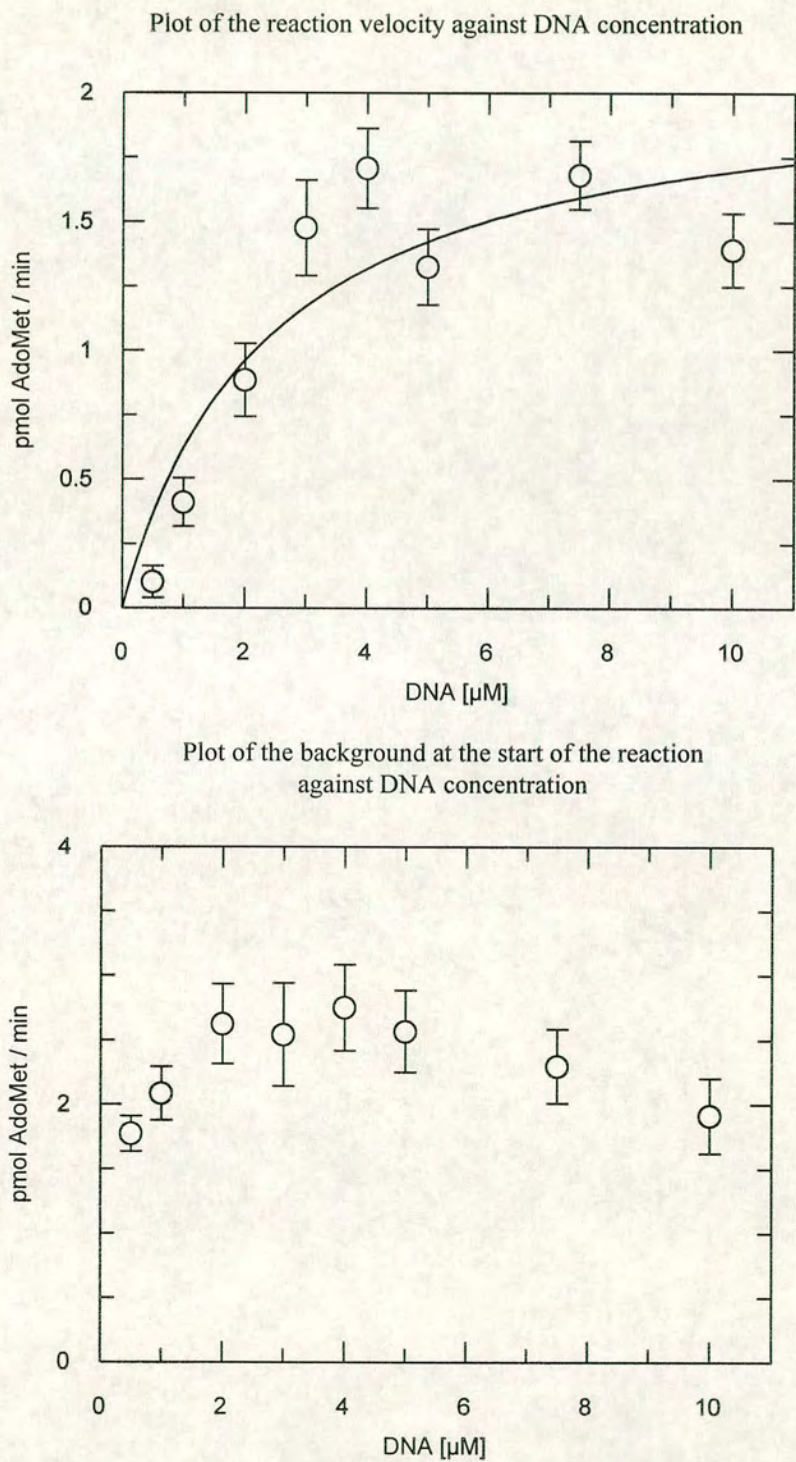
a) wild-type on unmethylated DNA





**Fig. 4.3.1. *In vitro* Assay of *M.Eco*KI methyltransferase activity by fluorographic Detection of [methyl-<sup>3</sup>H] Transfer - Dependence of the methylation reaction on DNA concentration.**

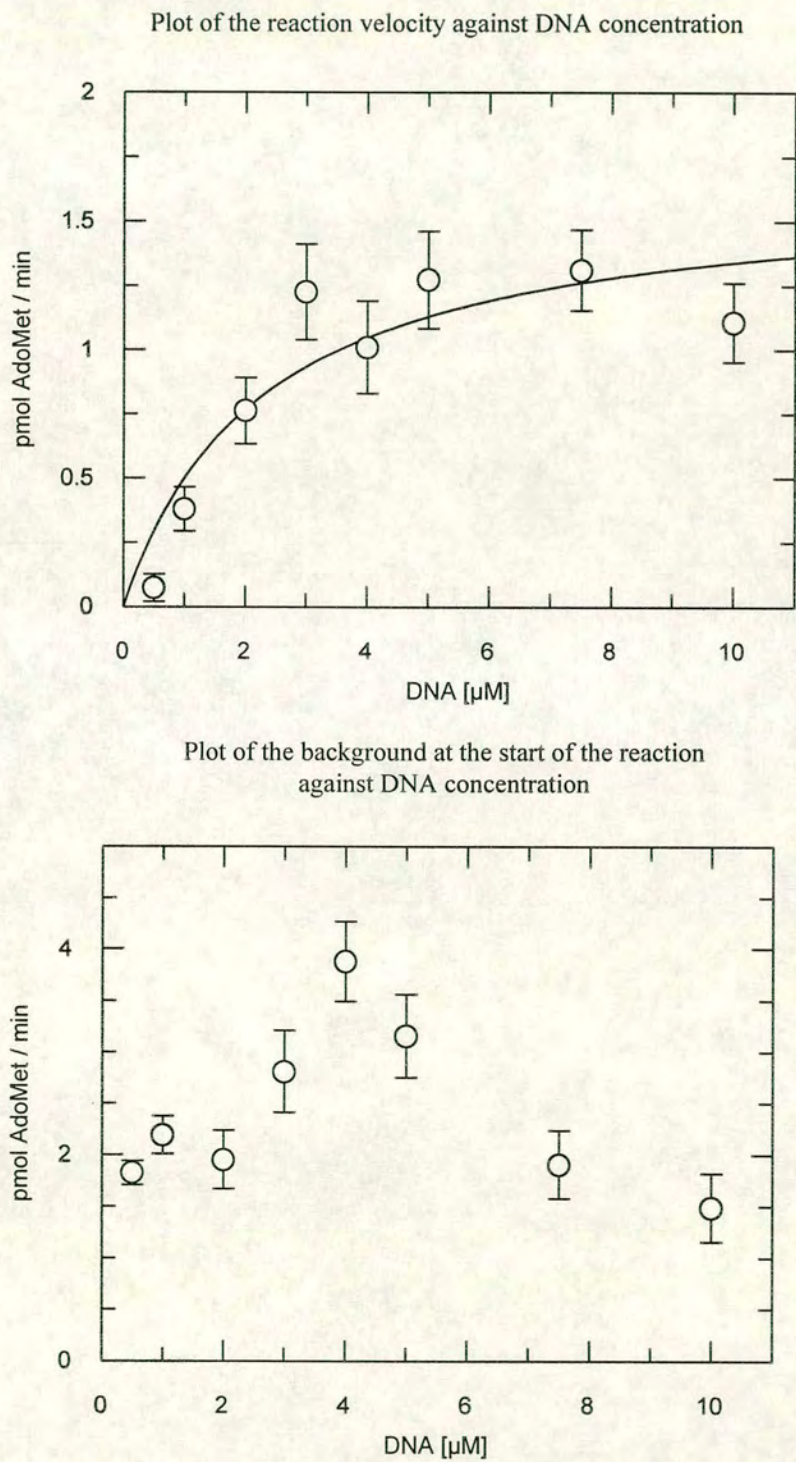
b) wild-type on hemi-methylated DNA (methylated in the upper strand)





**Fig. 4.3.1. *In vitro* Assay of M.EcoKI methyltransferase activity by fluorographic Detection of [methyl-<sup>3</sup>H] Transfer - Dependence of the methylation reaction on DNA concentration.**

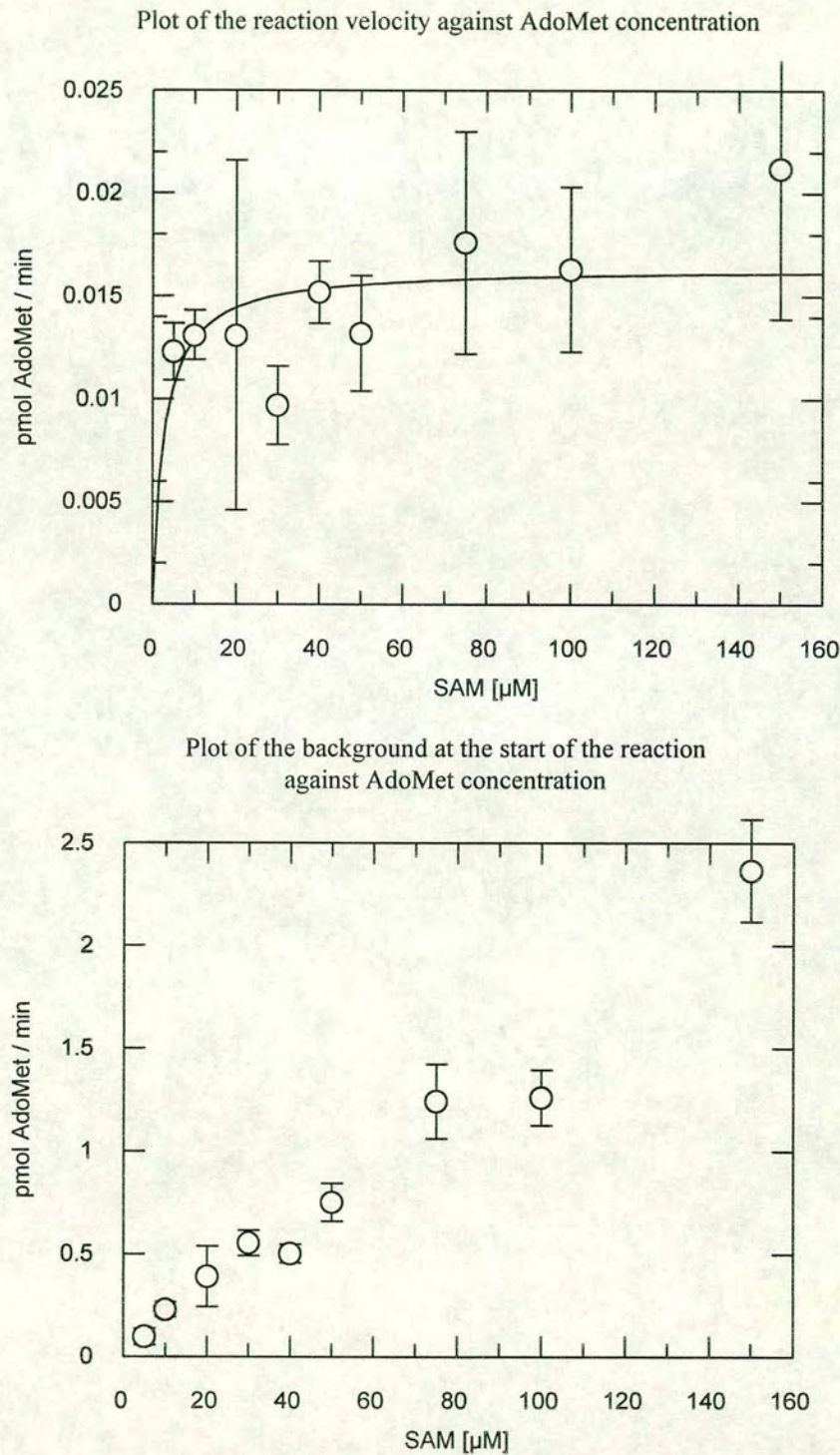
c) wild-type on hemi-methylated DNA (methylated in the lower strand))





**Fig. 4.3.2. *In vitro* Assay of *M.Eco*KI methyltransferase activity by fluorographic Detection of [methyl-<sup>3</sup>H] Transfer - Dependence of the methylation reaction on AdoMet concentration.**

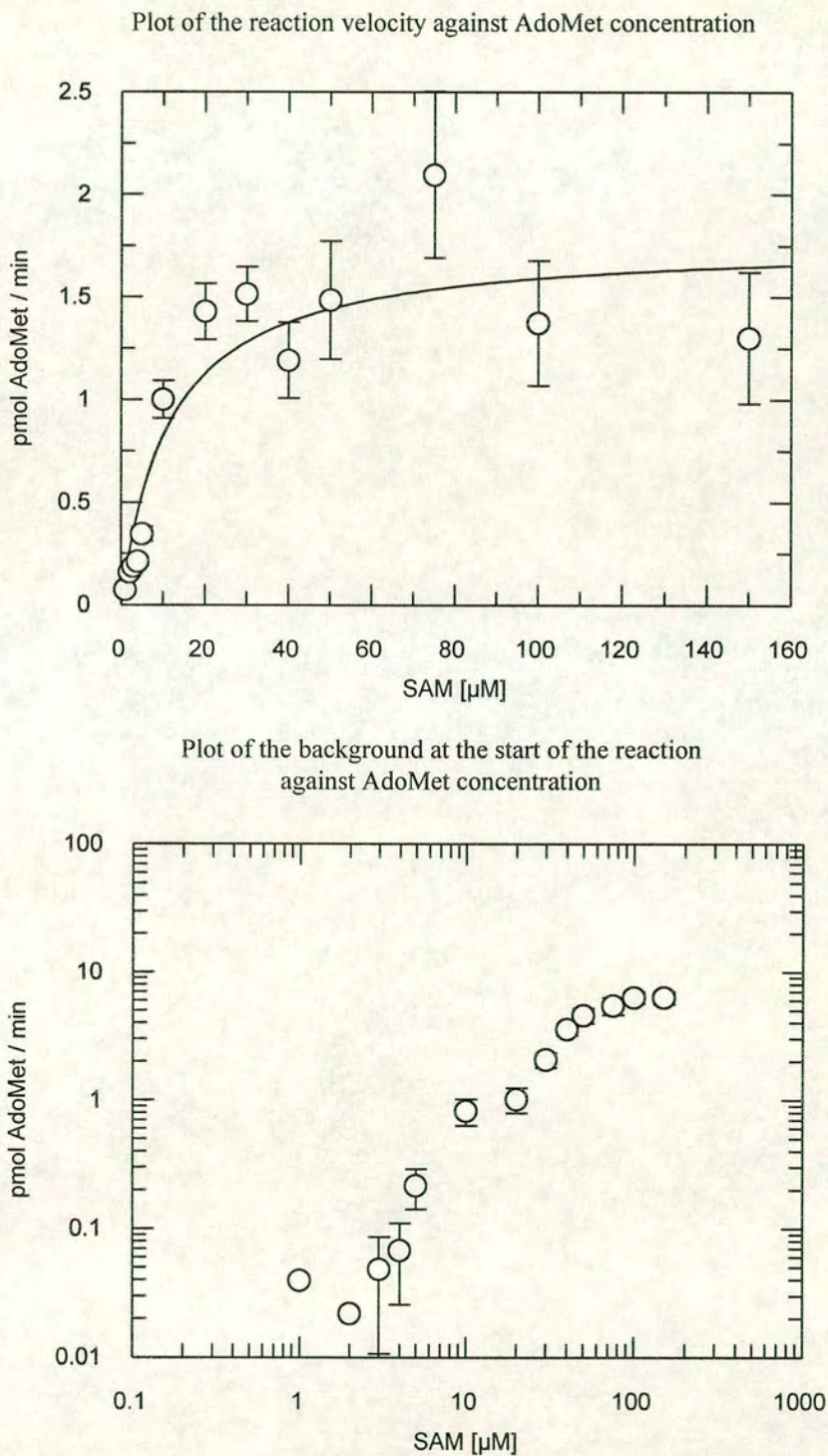
a) wild-type on unmethylated DNA





**Fig. 4.3.2. *In vitro* Assay of *M.Eco* KI methyltransferase activity by fluorographic Detection of [methyl-<sup>3</sup>H] Transfer - Dependence of the methylation reaction on AdoMet concentration.**

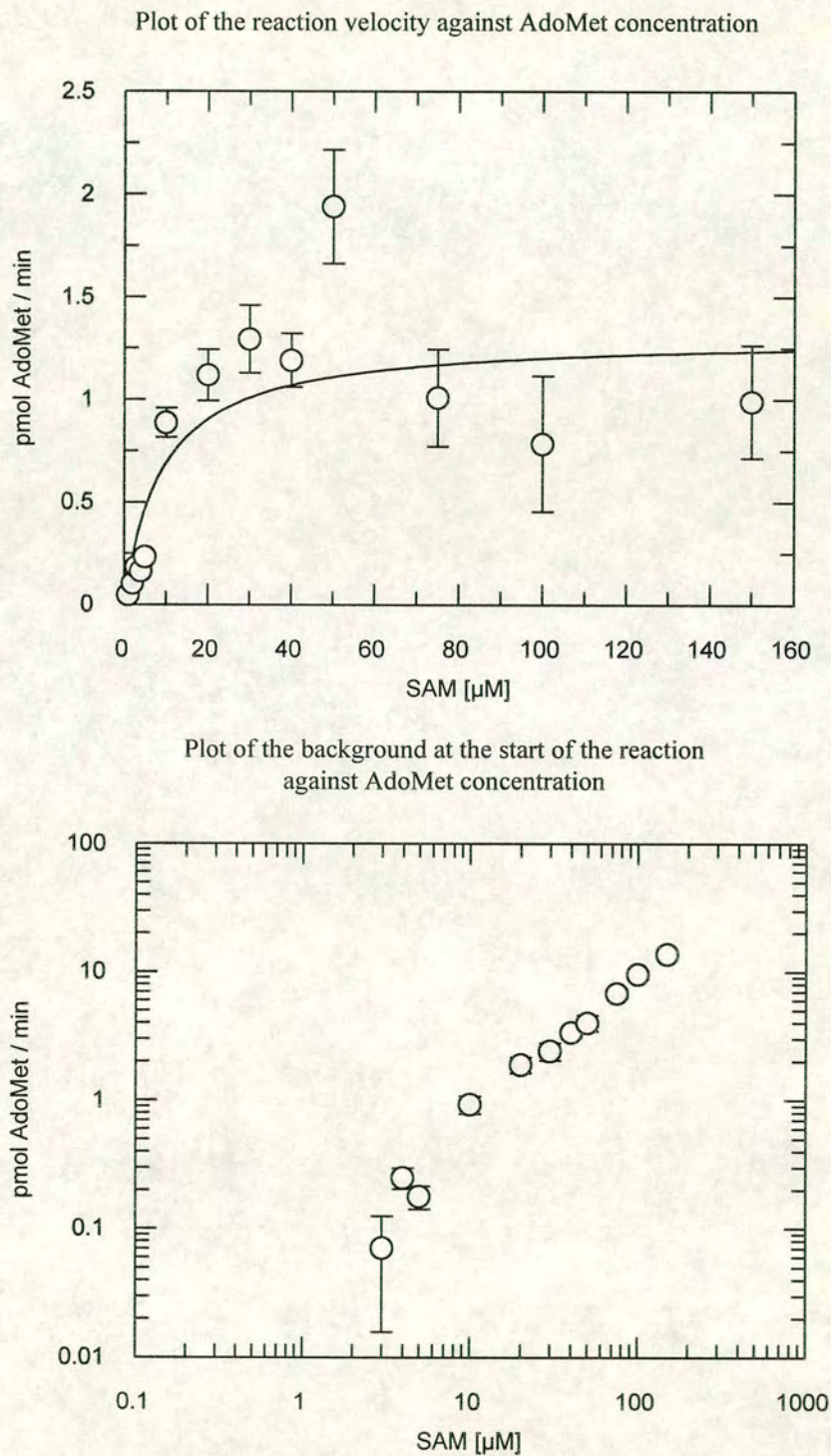
b) wild-type on hemi-methylated DNA (methylated in the upper strand)





**Fig. 4.3.2. *In vitro* Assay of *M.EcoKI* methyltransferase activity by fluorographic Detection of [methyl-<sup>3</sup>H] Transfer - Dependence of the methylation reaction on AdoMet concentration.**

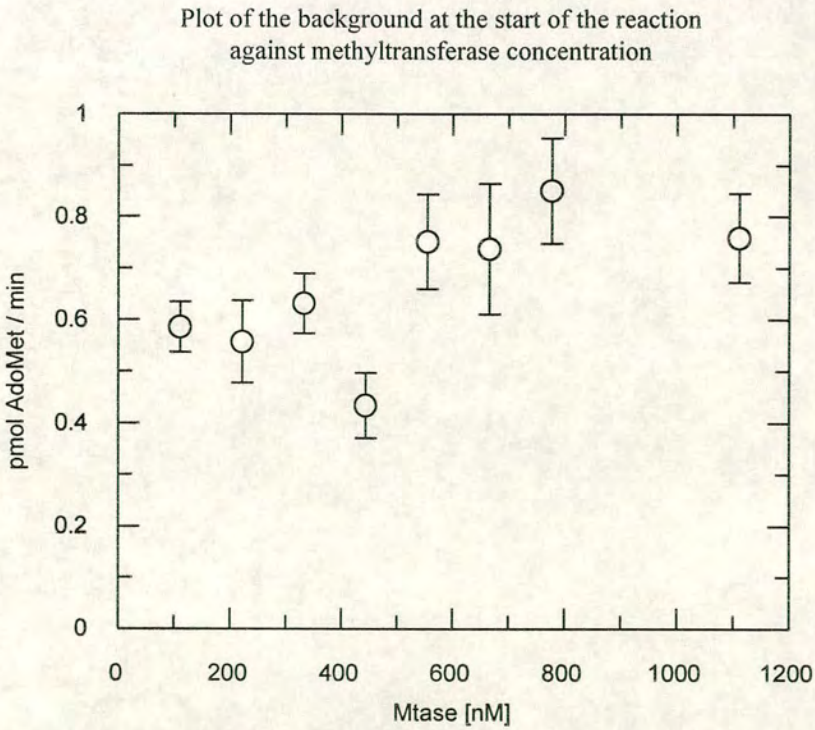
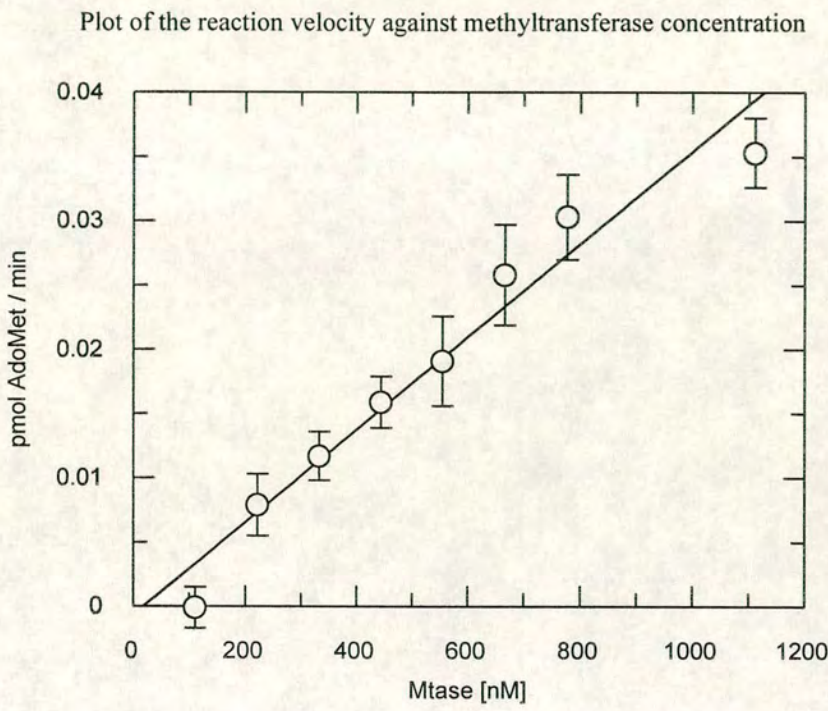
c) wild-type on hemi-methylated DNA (methylated in the lower strand))





**Fig. 4.3.3. *In vitro* Assay of M.EcoKI methyltransferase activity by fluorographic Detection of [methyl-<sup>3</sup>H] Transfer - Dependence of the methylation reaction on Methyltransferase concentration.**

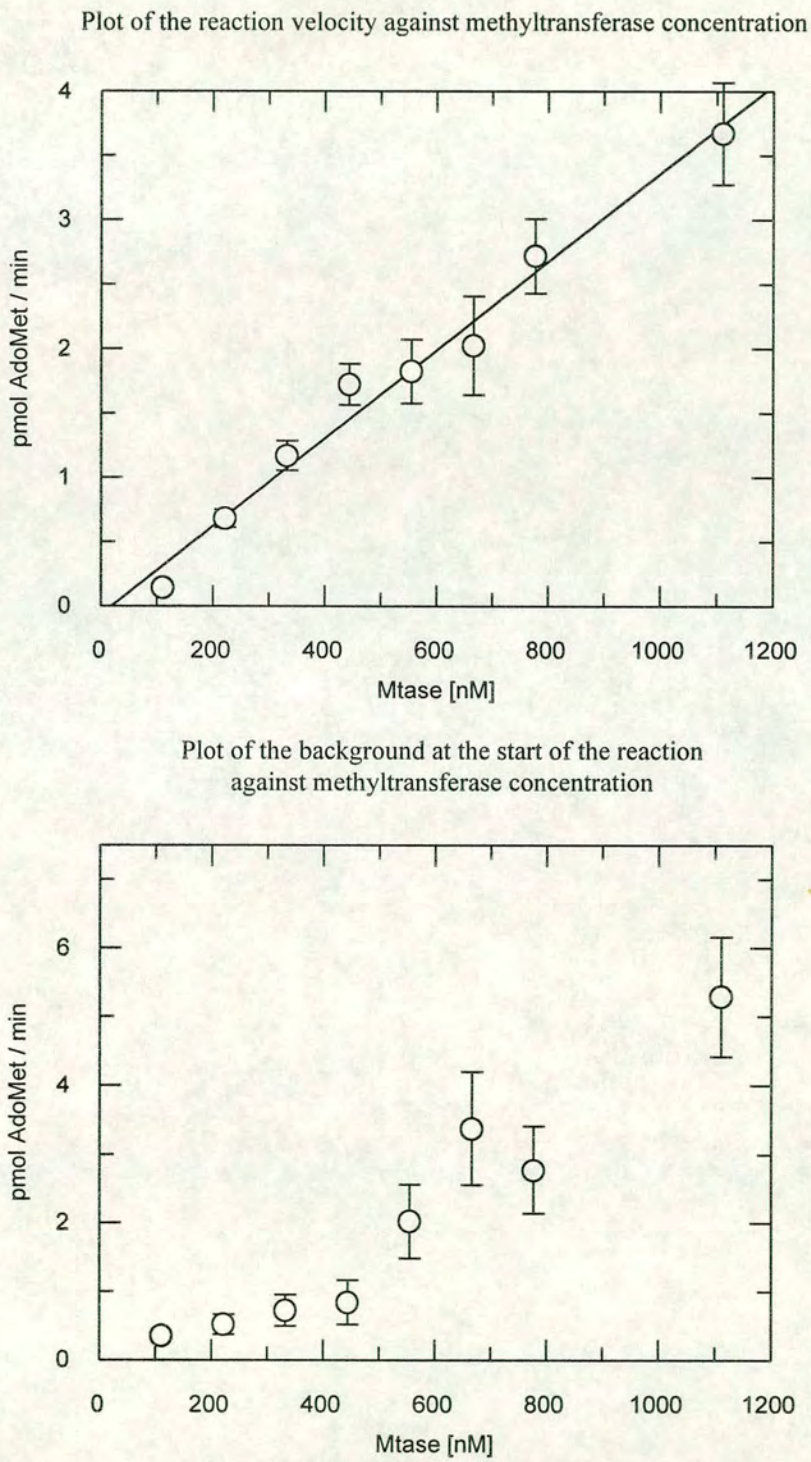
a) wild-type on unmethylated DNA





**Fig. 4.3.3. *In vitro* Assay of M.EcoKI methyltransferase activity by fluorographic Detection of [methyl-<sup>3</sup>H] Transfer - Dependence of the methylation reaction on Methyltransferase concentration.**

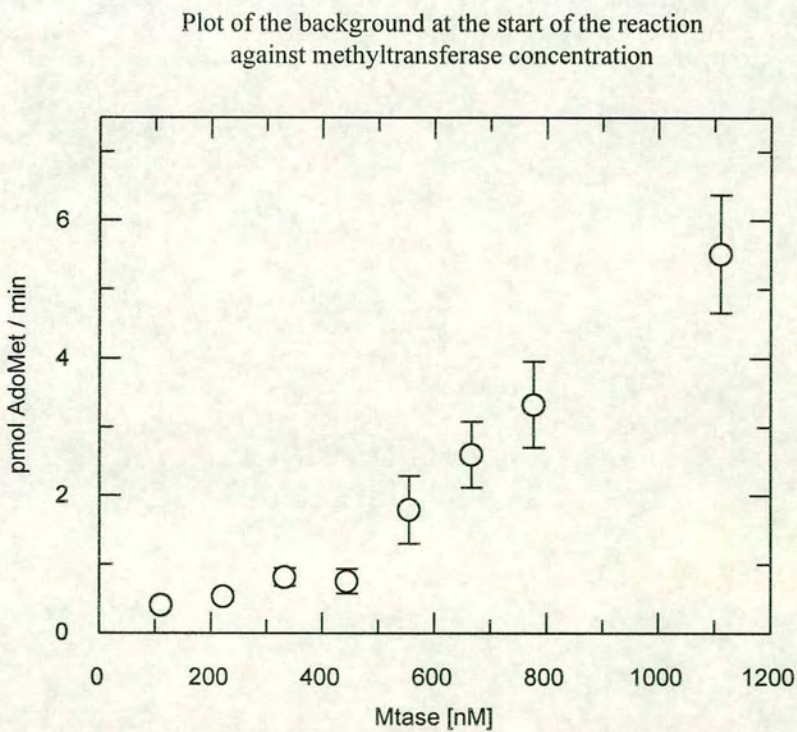
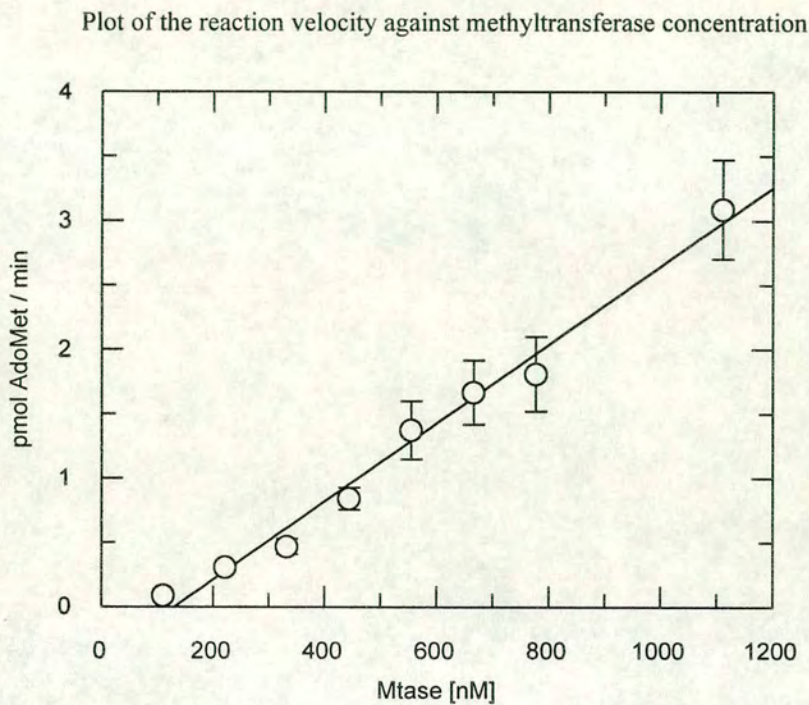
b) wild-type on hemi-methylated DNA (methylated in the upper strand)





**Fig. 4.3.3. *In vitro* Assay of M.EcoKI methyltransferase activity by fluorographic Detection of [methyl-<sup>3</sup>H] Transfer - Dependence of the methylation reaction on Methyltransferase concentration.**

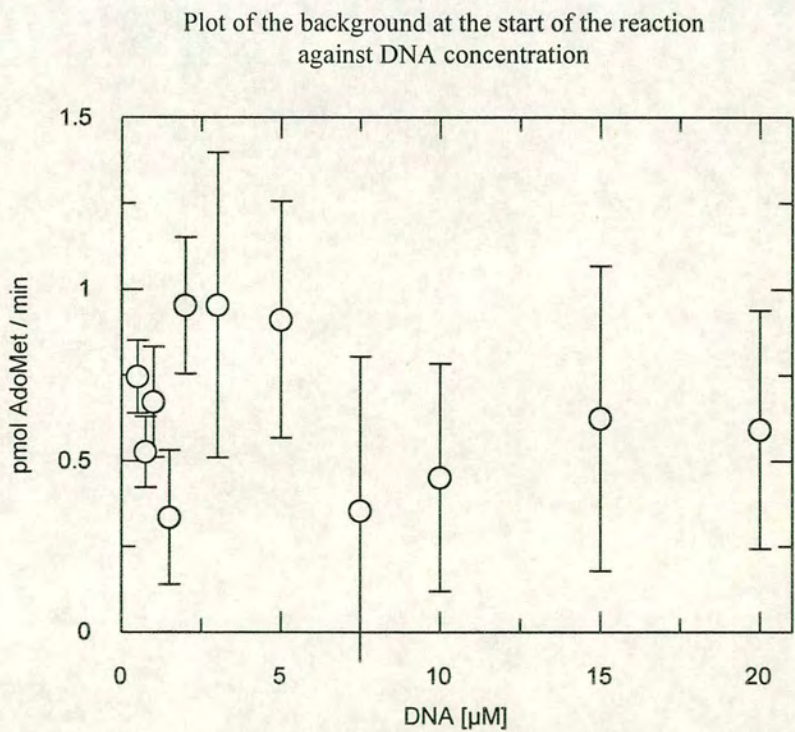
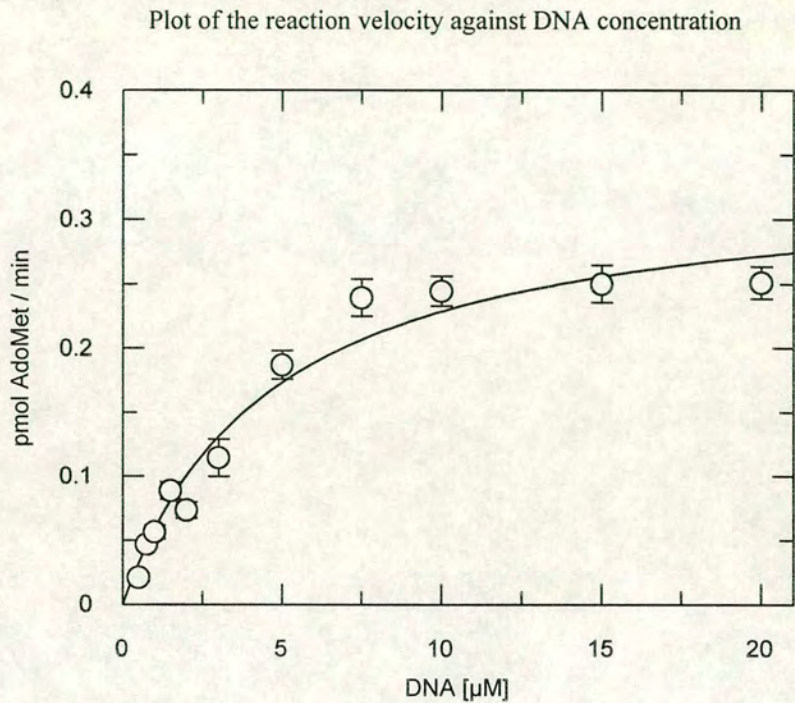
c) wild-type on hemi-methylated DNA (methylated in the lower strand))





**Fig. 4.4.1. *In vitro* Assay of *M.Eco*KI methyltransferase activity by fluorographic Detection of [methyl-<sup>3</sup>H] Transfer - Dependence of the methylation reaction on DNA concentration.**

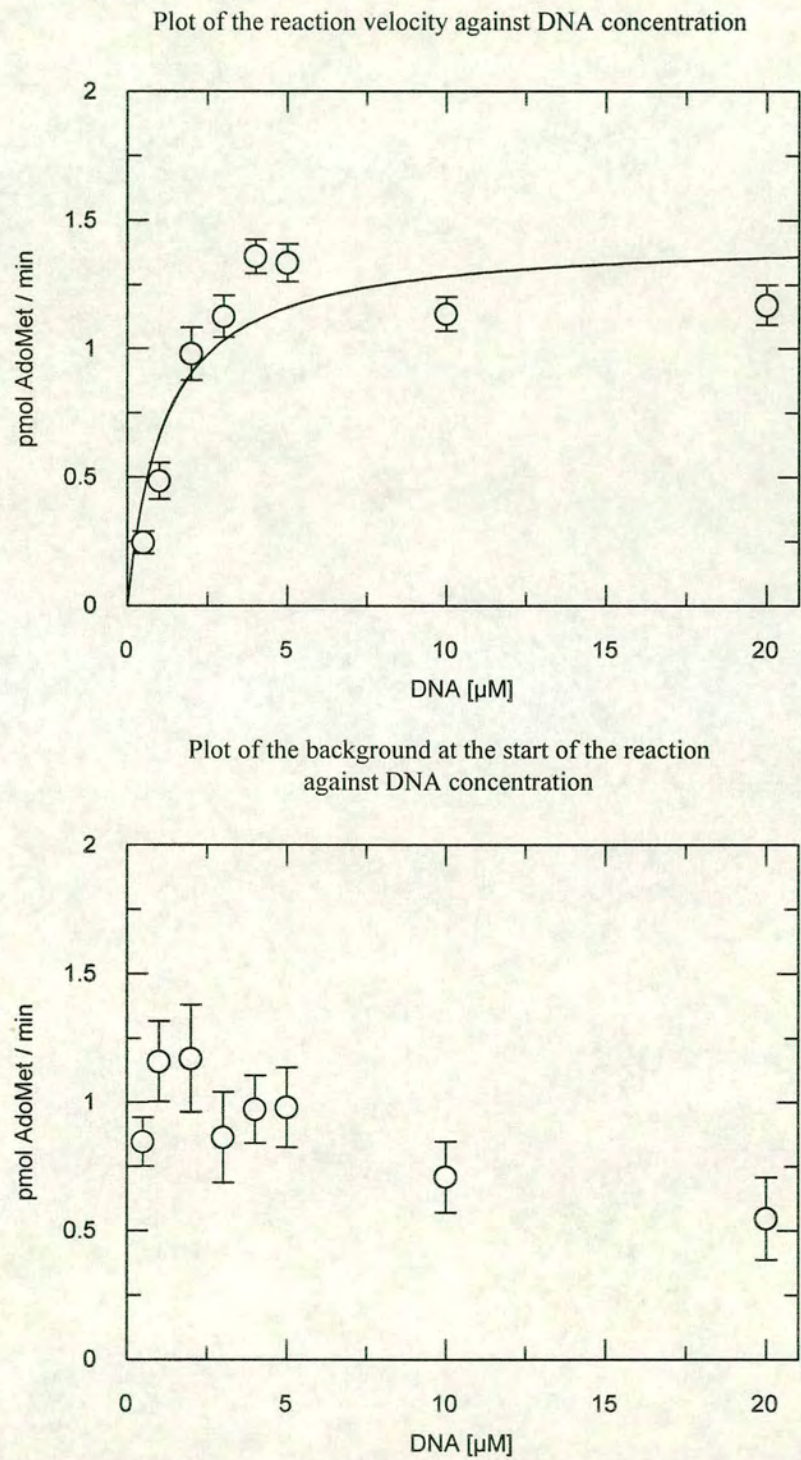
a) mutant L113Q on unmethylated DNA





**Fig. 4.4.1. *In vitro* Assay of M.EcoKI methyltransferase activity by fluorographic Detection of [methyl-<sup>3</sup>H] Transfer - Dependence of the methylation reaction on DNA concentration.**

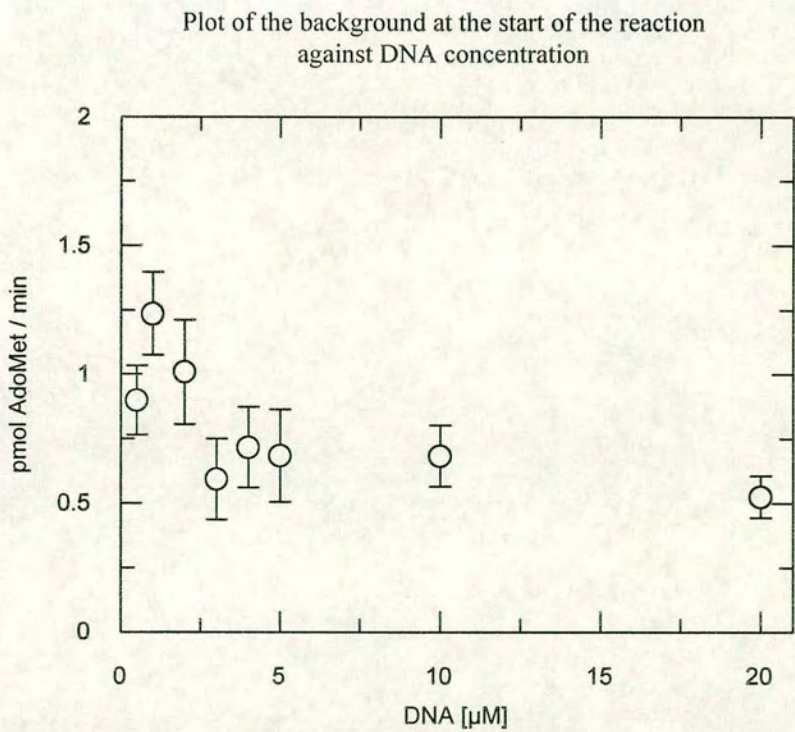
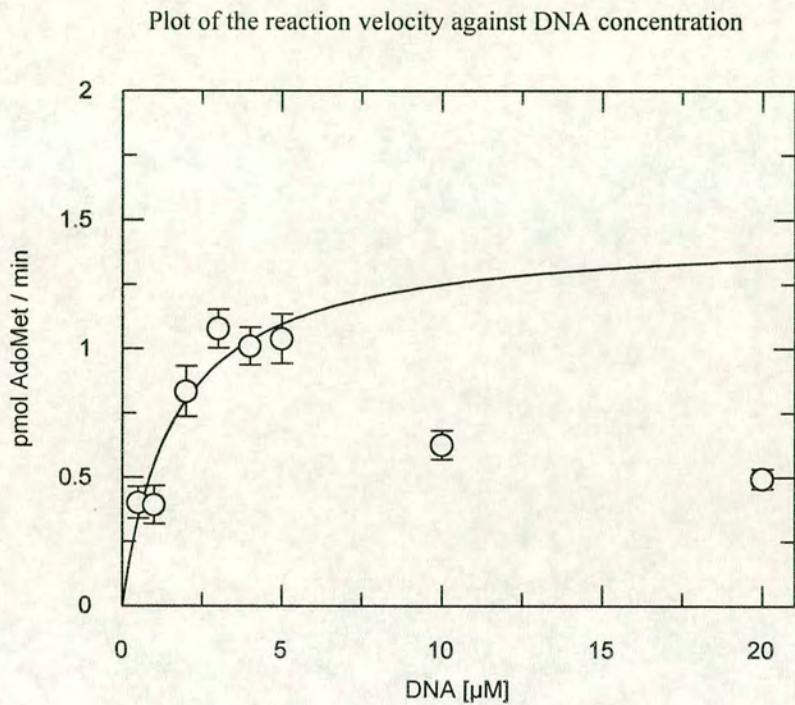
b) mutant L113Q on hemi-methylated DNA (methylated in the upper strand)





**Fig. 4.4.1. *In vitro* Assay of *M.Eco*KI methyltransferase activity by fluorographic Detection of [methyl-<sup>3</sup>H] Transfer - Dependence of the methylation reaction on DNA concentration.**

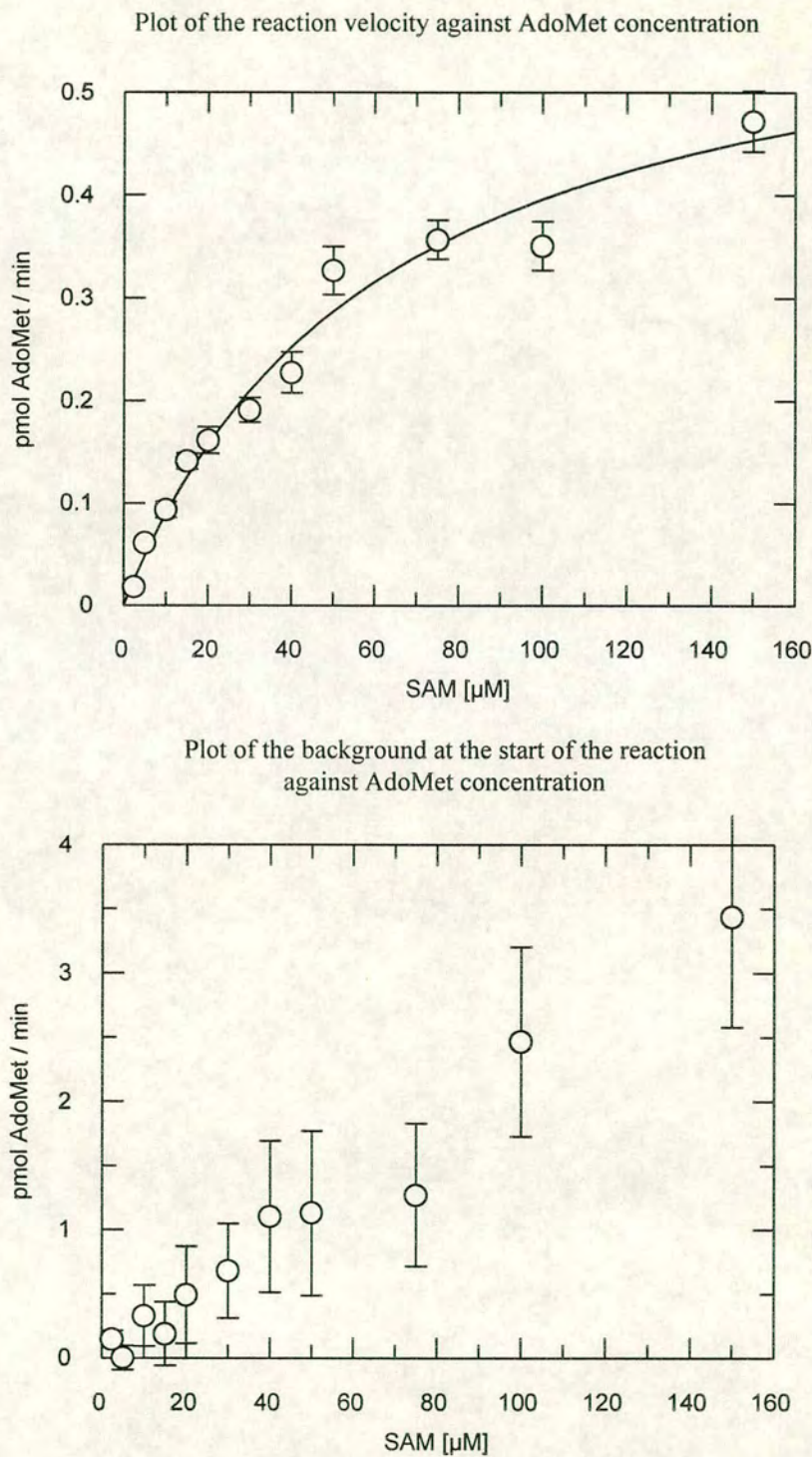
c) mutant L113Q on hemi-methylated DNA (methylated in the lower strand))





**Fig. 4.4.2. *In vitro* Assay of M.EcoKI methyltransferase activity by fluorographic Detection of [methyl-<sup>3</sup>H] Transfer - Dependence of the methylation reaction on AdoMet concentration.**

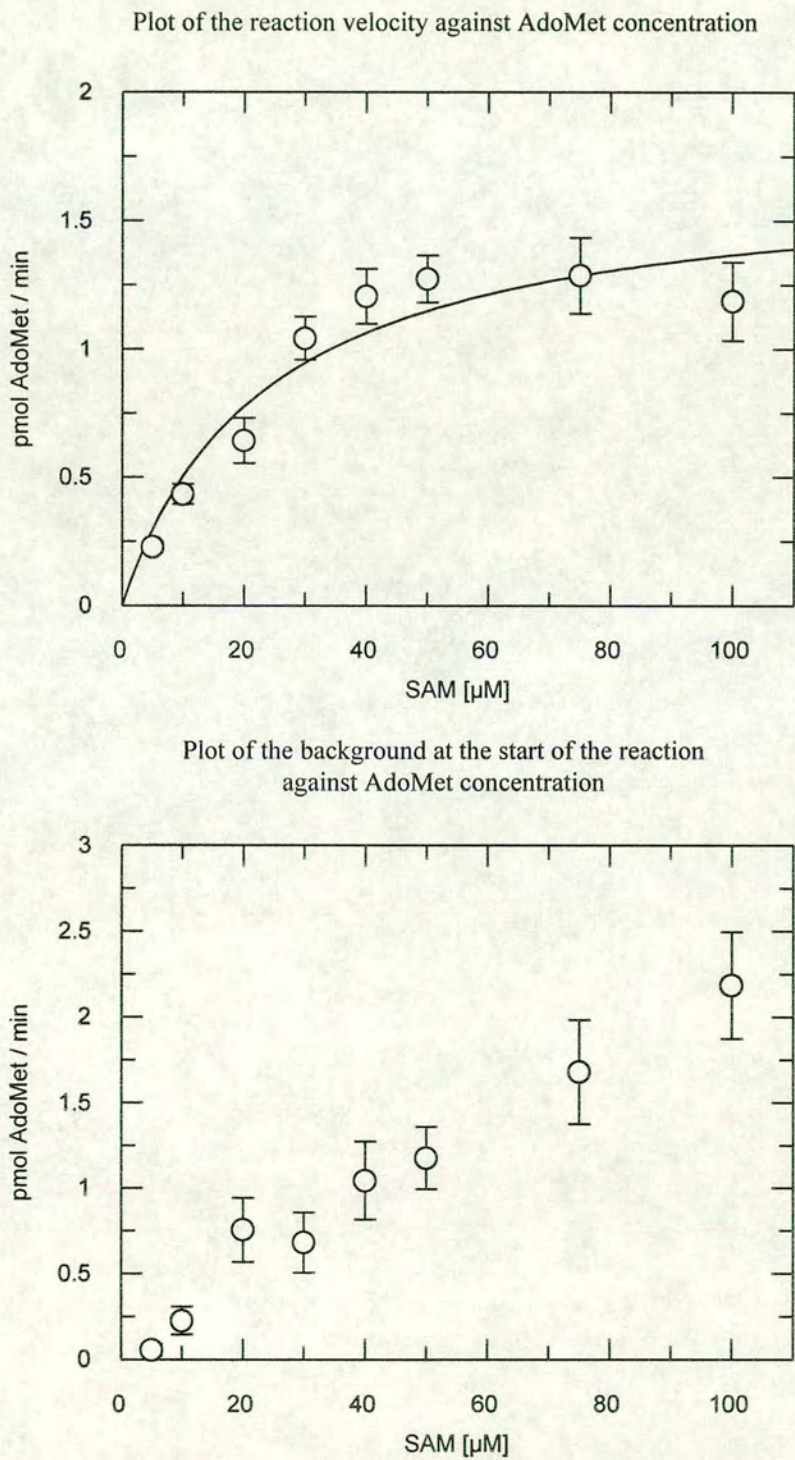
a) mutant L113Q on unmethylated DNA





**Fig. 4.4.2. *In vitro* Assay of *M.Eco*KI methyltransferase activity by fluorographic Detection of [methyl-<sup>3</sup>H] Transfer - Dependence of the methylation reaction on AdoMet concentration.**

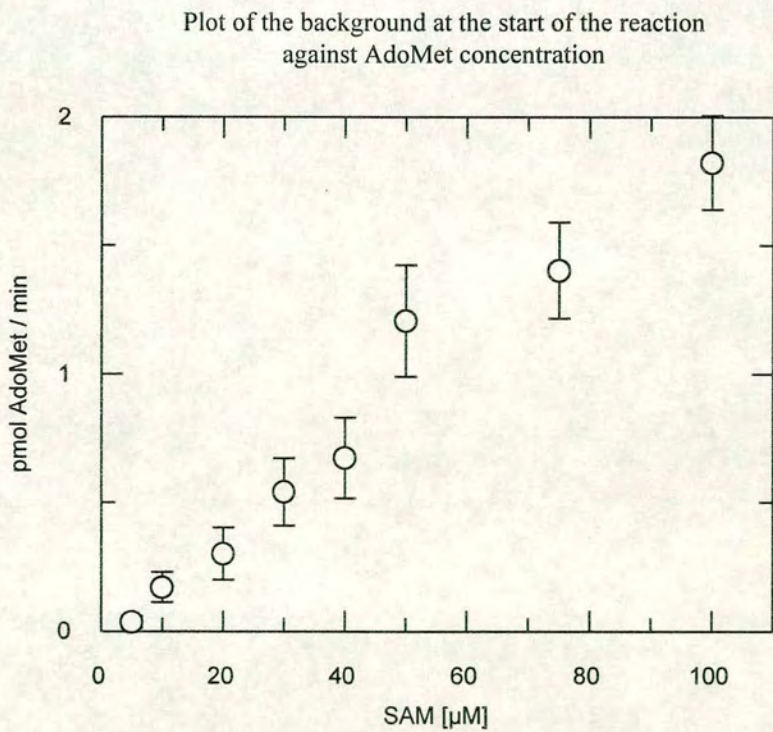
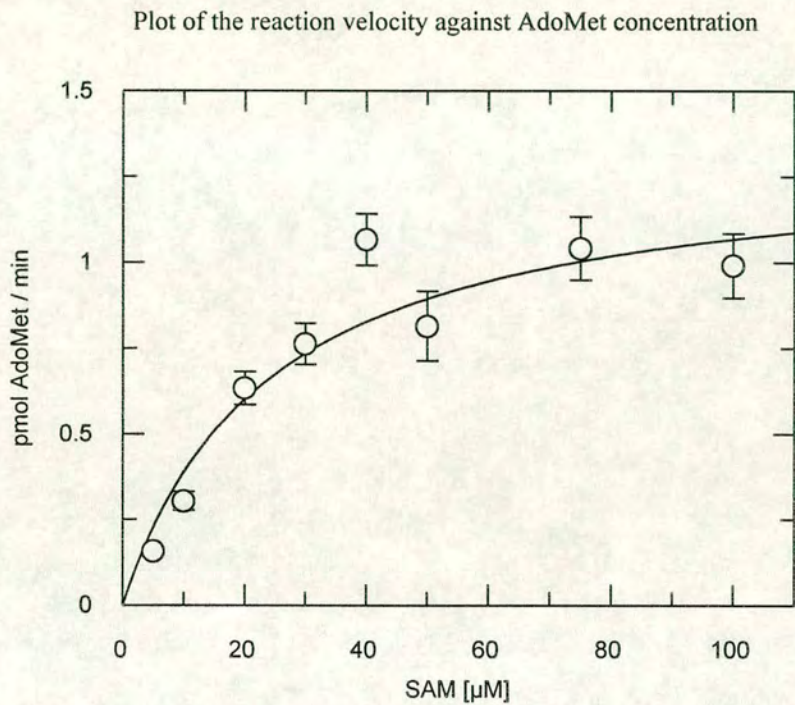
b) mutant L113Q on hemi-methylated DNA (methylated in the upper strand)





**Fig. 4.4.2. *In vitro* Assay of M.EcoKI methyltransferase activity by fluorographic Detection of [methyl-<sup>3</sup>H] Transfer - Dependence of the methylation reaction on AdoMet concentration.**

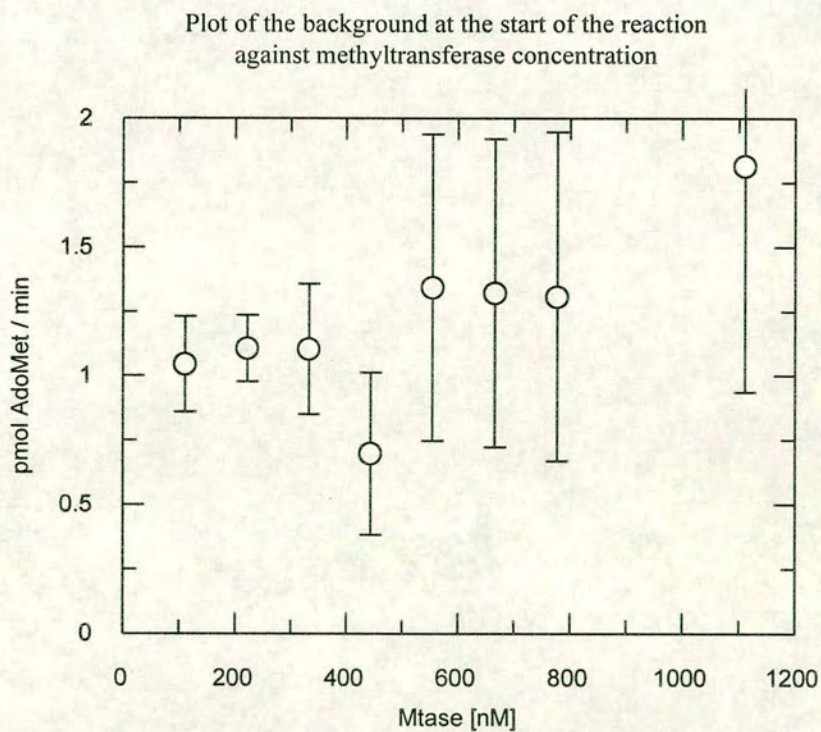
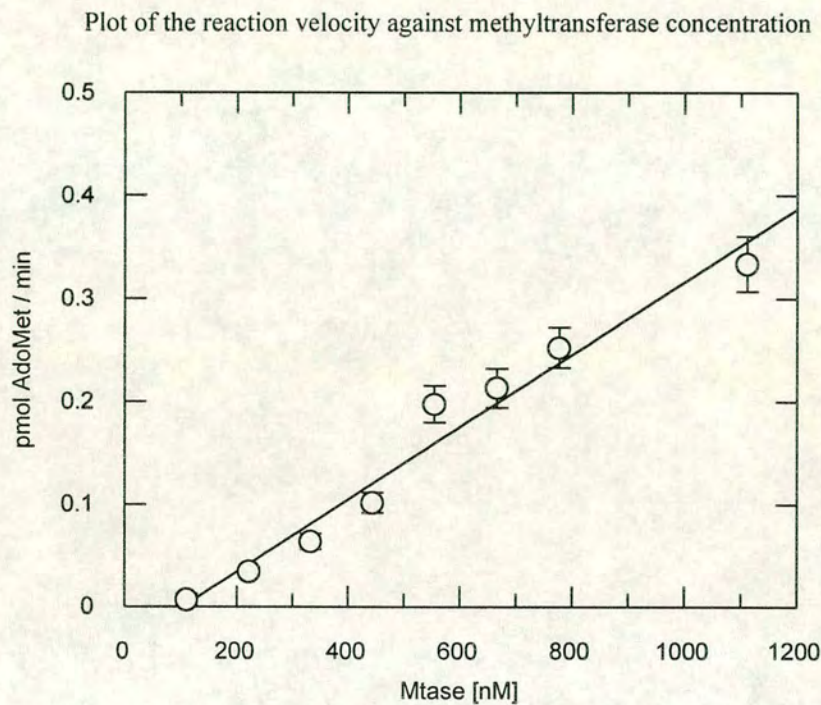
c) mutant L113Q on hemi-methylated DNA (methylated in the lower strand))





**Fig. 4.4.3. *In vitro* Assay of M.EcoKI methyltransferase activity by fluorographic Detection of [methyl-<sup>3</sup>H] Transfer - Dependence of the methylation reaction on Methyltransferase concentration.**

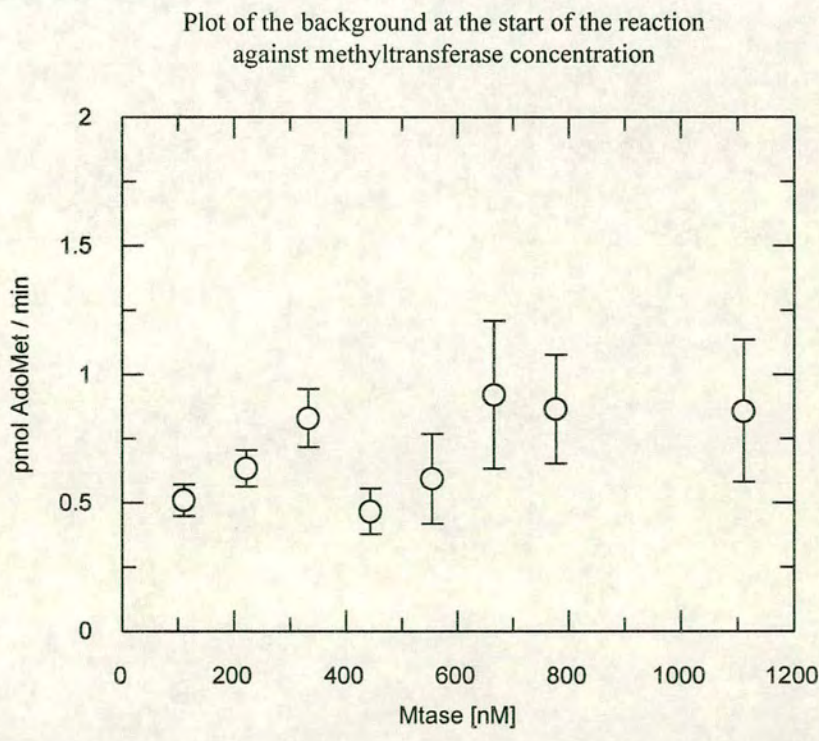
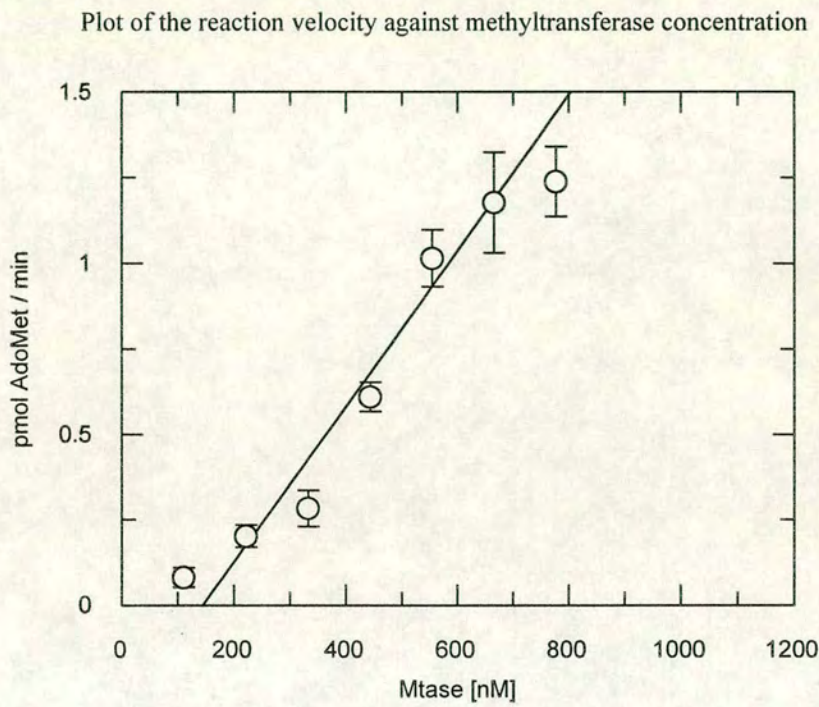
a) mutant L113Q on unmethylated DNA





**Fig. 4.4.3. *In vitro* Assay of M.EcoKI methyltransferase activity by fluorographic Detection of [methyl-<sup>3</sup>H] Transfer - Dependence of the methylation reaction on Methyltransferase concentration.**

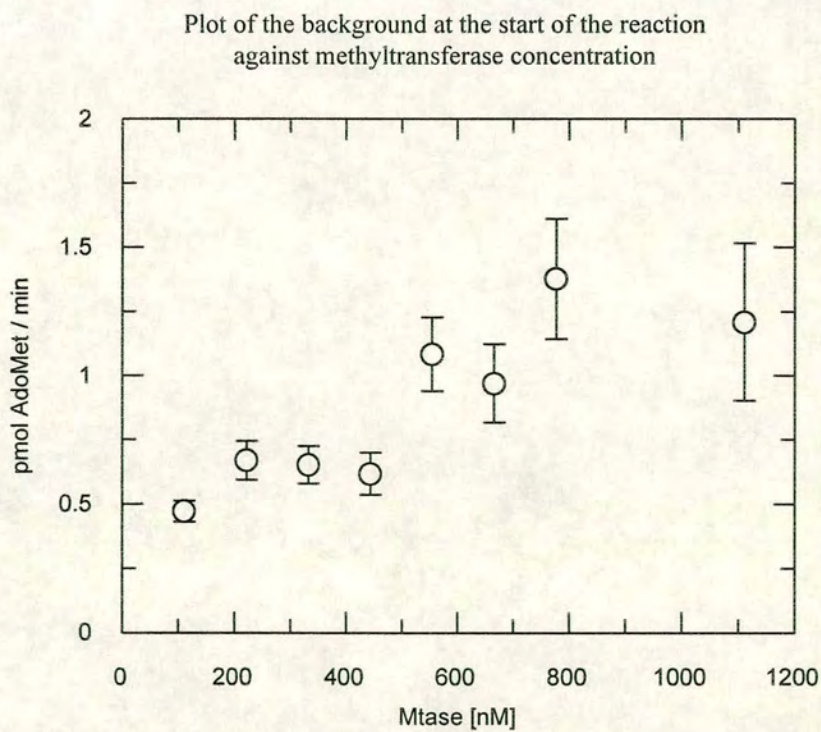
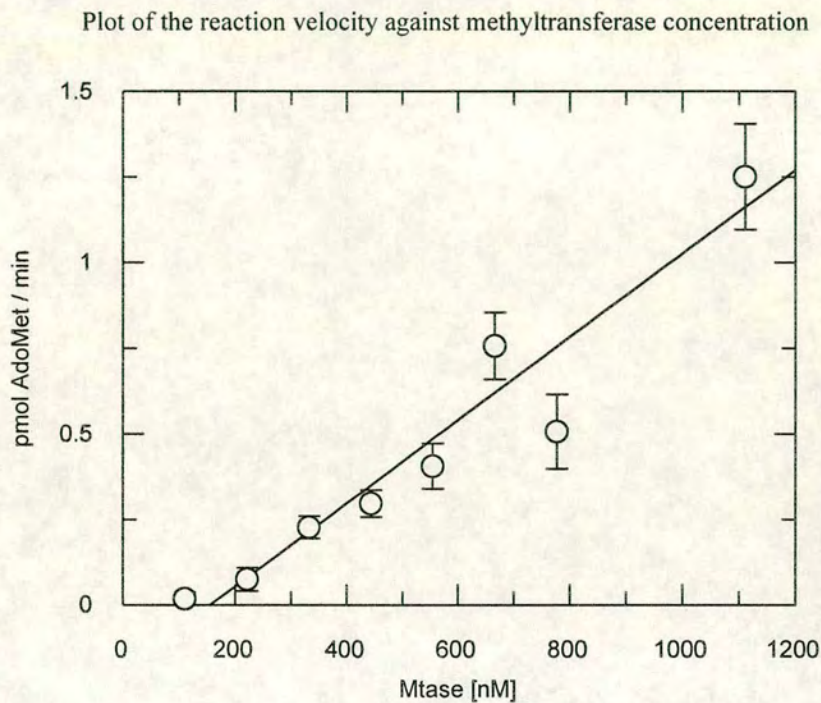
b) mutant L113Q on hemi-methylated DNA (methylated in the upper strand)





**Fig. 4.4.3. *In vitro* Assay of M.EcoKI methyltransferase activity by fluorographic Detection of [methyl-<sup>3</sup>H] Transfer - Dependence of the methylation reaction on Methyltransferase concentration.**

c) mutant L113Q on hemi-methylated DNA (methylated in the lower strand))





**Table 4.3: Kinetic constants derived from *in vitro* assays of the M.EcoKI methyltransferase activity**

Wild-Type M <sub>2</sub> S <sub>1</sub>	unmethylated DNA	hemi-methylated in upper strand	hemi-methylated in lower strand
<b>[DNA]</b>			
$v_{\max}$ [μM sec <sup>-1</sup> ]	(1.33 ± 0.13) * 10 <sup>-4</sup>	(1.75 ± 0.36) * 10 <sup>-2</sup>	(1.37 ± 0.24) * 10 <sup>-2</sup>
$K_m$ [M]	(0.88 ± 0.38) * 10 <sup>-6</sup>	(2.36 ± 1.38) * 10 <sup>-6</sup>	(2.28 ± 1.14) * 10 <sup>-6</sup>
$k_{cat}$ [sec <sup>-1</sup> ]	(2.40 ± 0.24) * 10 <sup>-4</sup>	(3.15 ± 0.65) * 10 <sup>-2</sup>	(2.47 ± 0.43) * 10 <sup>-2</sup>
$\frac{k_{cat}}{K_m}$ [M <sup>-1</sup> sec <sup>-1</sup> ]	(2.74 ± 0.63) * 10 <sup>2</sup>	(1.33 ± 0.47) * 10 <sup>4</sup>	(1.08 ± 0.38) * 10 <sup>4</sup>
<b>[AdoMet]</b>			
$v_{\max}$ [μM sec <sup>-1</sup> ]	(1.37 ± 0.13) * 10 <sup>-4</sup>	(1.48 ± 0.16) * 10 <sup>-2</sup>	(1.09 ± 0.16) * 10 <sup>-2</sup>
$K_m$ [M]	(2.71 ± 1.99) * 10 <sup>-6</sup>	(11.49 ± 4.61) * 10 <sup>-6</sup>	(8.91 ± 5.04) * 10 <sup>-6</sup>
$k_{cat}$ [sec <sup>-1</sup> ]	(2.46 ± 0.23) * 10 <sup>-4</sup>	(2.66 ± 0.29) * 10 <sup>-2</sup>	(1.96 ± 0.29) * 10 <sup>-2</sup>
$\frac{k_{cat}}{K_m}$ [M <sup>-1</sup> sec <sup>-1</sup> ]	(0.91 ± 0.11) * 10 <sup>2</sup>	(2.32 ± 0.64) * 10 <sup>3</sup>	(2.20 ± 0.58) * 10 <sup>3</sup>
<b>[Mtase]</b>			
intercept with x-axis	19 ± 51 nM	19 ± 32 nM	133 ± 30 nM

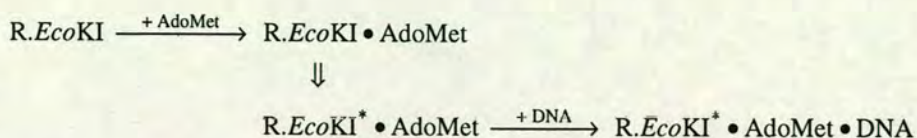


**Table 4.4: Kinetic constants derived from *in vitro* assays of the M.EcoKI methyltransferase activity**

L113Q M <sub>2</sub> S <sub>1</sub>	unmethylated DNA	hemi-methylated in upper strand	hemi-methylated in lower strand
[DNA]			
v <sub>max</sub> [μM sec <sup>-1</sup> ]	(2.79 ± 0.21) * 10 <sup>-3</sup>	(1.19 ± 0.14) * 10 <sup>-2</sup>	(1.21 ± 0.19) * 10 <sup>-2</sup>
K <sub>m</sub> [M]	(4.66 ± 0.89) * 10 <sup>-6</sup>	(1.17 ± 0.53) * 10 <sup>-6</sup>	(1.62 ± 0.69) * 10 <sup>-6</sup>
k <sub>cat</sub> [sec <sup>-1</sup> ]	(5.03 ± 0.37) * 10 <sup>-3</sup>	(2.15 ± 0.25) * 10 <sup>-2</sup>	(2.18 ± 0.34) * 10 <sup>-2</sup>
$\frac{k_{cat}}{K_m}$ [M <sup>-1</sup> sec <sup>-1</sup> ]	(1.08 ± 0.42) * 10 <sup>3</sup>	(1.85 ± 0.47) * 10 <sup>4</sup>	(1.35 ± 0.49) * 10 <sup>4</sup>
[AdoMet]			
v <sub>max</sub> [μM sec <sup>-1</sup> ]	(5.33 ± 0.45) * 10 <sup>-3</sup>	(1.41 ± 0.17) * 10 <sup>-2</sup>	(1.10 ± 0.14) * 10 <sup>-2</sup>
K <sub>m</sub> [M]	(61.33 ± 10.63) * 10 <sup>-6</sup>	(23.48 ± 7.97) * 10 <sup>-6</sup>	(24.02 ± 8.19) * 10 <sup>-6</sup>
k <sub>cat</sub> [sec <sup>-1</sup> ]	(9.60 ± 0.80) * 10 <sup>-3</sup>	(2.54 ± 0.31) * 10 <sup>-2</sup>	(1.99 ± 0.24) * 10 <sup>-2</sup>
$\frac{k_{cat}}{K_m}$ [M <sup>-1</sup> sec <sup>-1</sup> ]	(1.56 ± 0.75) * 10 <sup>2</sup>	(1.08 ± 0.39) * 10 <sup>3</sup>	(0.83 ± 0.30) * 10 <sup>3</sup>
[Mtase]			
intercept with x-axis	103 ± 43 nM	143 ± 44 nM	158 ± 68 nM



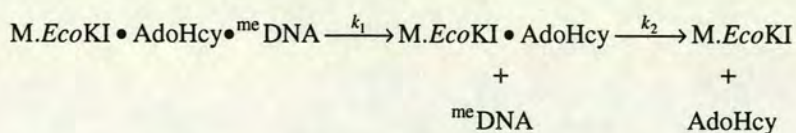
Earlier studies on the enzymatic mechanism of the *Eco*KI system were done using the complete endonuclease complex, and established the following reaction scheme:



in which the slow transition from *R.Eco*KI to *R.Eco*KI\* was identified as the rate limiting step. Formation of a filter binding complex by *R.Eco*KI with unmodified  $\lambda$  DNA showed a sigmoid dependence on AdoMet concentration, thus identifying AdoMet as a positive allosteric effector in the nucleolytic reaction with an apparent  $K_m$  of  $\approx 0.3 \mu\text{M}$ .

The availability of large amounts of purified *Eco*KI methyltransferase made investigations into the enzymatic process of the  $\text{M}_2\text{S}_1$  complex possible. Despite the large amount of enzyme used in the assays, the formation of the methylated DNA product as followed by scintillation counting was linear with time over an appreciable period, making accurate measurements of initial rates possible. An exception to this was seen at very high AdoMet concentrations (75 - 150  $\mu\text{M}$ ), where the progress curve soon deviated from linearity, making the values obtained less reliable. It seems likely that the DNA concentration became limiting in these cases, thereby resulting in the true velocity of the reaction being underestimated.

Linear regression analysis was applied to all progress curves and both the slope and the intercept with the y-axis of the resulting line recorded. An increase in DNA concentration has had little effect on the extrapolated intercept. However an increase in AdoMet concentration resulted in a linear increase in the extrapolated intercept, suggesting an initial burst of product formation. This type of situation occurs when an enzyme-bound intermediate accumulates during the reaction. The first mole of substrate rapidly reacts with the enzyme to form stoichiometric amounts of the enzyme-bound intermediate and product, whilst the subsequent reaction is slow since it depends on the slow breakdown of the intermediate to release free enzyme. For the methylation reaction of the *Eco*KI methyltransferase this could be represented as:





Hypothetically if  $k_1$  is fast and  $k_2$  so slow as to be negligible, then an initial burst of product formation is seen in the reaction, but no further production of methylated DNA occurs. However if  $k_1$  is much faster than  $k_2$ , but  $k_2$  is not negligible, then the initial burst of formation of methylated DNA is followed by a progressive increase as the intermediate turns over. The continued formation of methylated DNA would then depend on how fast the free enzyme is regenerated, and  $k_2$  would be rate limiting.

However there are two serious objections against this interpretation of the data. Firstly the size of the burst is limited by the amount of enzyme, and therefore cannot be greater than the amount of enzyme in the reaction.

$$\text{burst size} = [E_0] * \left( \frac{k_1}{k_1 + k_2} \right)^2$$

The observed "bursts" however exceed the limit set by the amount of enzyme in the reaction by up to several times. Secondly the reaction proceeds in a linear fashion even after large "bursts of activity" would have been expected to seriously deplete the supply of DNA. It can therefore be concluded that the apparent "bursts" that go with an increase in the amount of AdoMet present in the sample are most likely artefacts caused by cofactor being retained on the filter paper despite several washing steps.

Disregarding "bursts" and the associated burst kinetics, the methyltransferase seems to obey Michaelis-Menten kinetics for both varying DNA and AdoMet concentrations. Thus the Michaelis-Menten equation

$$v = v_{\max} * \frac{[S]}{K_m + [S]} = [E_0] * k_{cat} * \frac{[S]}{K_m + [S]}$$

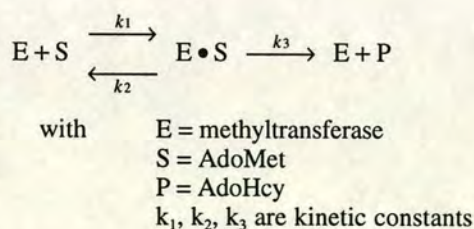
was fitted to the data set and the kinetic values derived from the fit. With the total enzyme concentration  $E_0$  known,  $k_{cat}$  and the ratio  $k_{cat}/K_m$  could also be calculated. For more complex reactions like the methylation by Type I methyltransferases,  $k_{cat}$  is a function of all the first-order rate constants, and cannot be greater than any one of the first-order rate constants on the forward reaction pathway. It therefore sets a lower limit on the value of the chemical rate constants. Because it also represents the maximum number of substrate molecules converted to products per active site per unit time, it is often also called the *turnover* number. Conversely  $k_{cat}/K_m$  cannot be greater than any second-order rate constant on the forward reaction pathway. It thus sets a lower limit on the rate constant for the association of enzyme and substrate, and can also be regarded as a measure for an enzymes catalytic efficiency.



Although unmodified DNA is a very good substrate in the filter-binding reaction of R.*Eco*KI, it makes a very poor substrate for the methylation reaction. Consequently the sigmoid dependence on AdoMet concentration seen in the filter-binding reaction ( $K_m = 0.3 \mu\text{M}$ ) could not be investigated for the wild-type enzyme since virtual no methylation occurs on unmodified DNA at low AdoMet concentrations.

The *in vitro* methylase activity results for the wild-type methyltransferase are in excellent agreement with previously reported *in vivo* data. Both hemi-methylated DNA substrates are methylated at the same rate, and about 100 times faster than the unmodified substrate. The  $K_m$  for both hemi-methylated DNA substrates is nearly identical, and similar to the  $K_m$  for the unmodified substrate. Equally the  $K_m$  values for AdoMet are nearly identical in the methylation reactions involving hemi-methylated substrate, and similar to the  $K_m$  for AdoMet in the methylation of unmodified substrate. That the  $K_m$  values for unmodified and hemi-methylated substrates are so similar while the respective velocities vary so much indicates that the enzyme distinguishes between unmodified and hemi-methylated substrate at the level of catalysis, not at the level of substrate binding.

It is noteworthy that the  $K_m$  for AdoMet in the methylation reaction is very similar to the  $K_d$  for AdoMet as determined by both the titration of an ANS/methyltransferase solution with AdoMet and by equilibrium dialysis (Powell *et al.*, 1993). Given a simplified scheme where



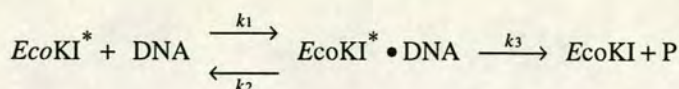
and accordingly where

$$K_m = \frac{k_2 + k_3}{k_1} \cong \frac{k_2}{k_1} = K_d \quad \text{only when } k_3 \text{ is much smaller than } k_2$$

it would indicate that the rate of the reaction is limited at a stage after formation of the enzyme-AdoMet complex.

The  $K_m$  values for the DNA substrate determined during the methylation reaction are about a thousand times higher than the  $K_d$  values determined by Powell *et al.* (1993). The same simplified scheme





shows that

$$K_m = \frac{k_2 + k_3}{k_1} \gg \frac{k_2}{k_1} = K_d$$

only when  $k_3$  is to be significantly larger than  $k_2$ .

The  $k_{\text{cat}}$  values of the wild-type methylase for AdoMet are virtually identical to its  $k_{\text{cat}}$  values for DNA, as would be expected for a reaction that turns over the same numbers of each substrate, *e.g.* one AdoMet is used per DNA molecule. This is true irrespective of the methylation state of the DNA substrate, and the wild-type enzyme therefore methylates just once per catalytic cycle. Methylation of unmethylated DNA must therefore occur in two separate methylation reactions.

The same is true for the L113Q mutant regarding hemi-methylated DNA. Its  $k_{\text{cat}}$  values are furthermore nearly identical to the values of the wild-type enzyme, and it therefore methylates hemi-methylated DNA in the same way as the wild-type enzyme despite its lower affinity for AdoMet. On unmodified DNA however the mutant behaves quite differently. Not only does the L113Q mutant methylate unmodified DNA about 20 times faster than the wild-type enzyme, it also seems to turn over twice as much AdoMet as DNA. This result can be interpreted in two ways. The first interpretation is that the mutant methylates both methylation target sites in one catalytic cycle and probably in a concerted reaction at the same time, thereby turning over two moles of AdoMet per mole DNA. However this is in stark contrast to the fact that the mutants bind AdoMet with lower affinity, and probably have only one AdoMet molecule bound under the conditions used. The second interpretation is that the wild-type binds AdoMet with high affinity, and is working under saturating AdoMet concentrations in both assays - in the one varying the DNA concentration which use 50  $\mu\text{M}$  AdoMet as well as in the one varying the AdoMet concentration from 1 to 150  $\mu\text{M}$  AdoMet. Consequently the same  $v_{\text{max}}$  is reached in both cases, and the same turnover number will be displayed. The mutant however binds AdoMet worse - it will not reach saturating levels in the assays varying the DNA concentration (thereby displaying a lower  $v_{\text{max}}$  and a lower turnover number), but will do so in the assays varying the AdoMet concentration (or at least reach higher saturation), thus displaying a higher  $v_{\text{max}}$  and a higher turnover number. This interpretation is fully compatible with the ANS titration results.



The low affinity of the mutant methyltransferase for AdoMet as seen in the ANS titration studies is reflected in the higher  $K_m$  values for AdoMet. That this higher  $K_m$  for AdoMet does not lead to a change in  $v_{max}$  for the methylation of hemi-methylated DNA indicates that the weaker binding of AdoMet by the L113Q mutant does not influence the methylation reaction. As a consequence the lower affinity for AdoMet can not be the cause for the observed behaviour of the mutant methyltransferase, but is just an effect of the same.

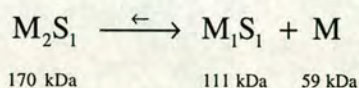
The rate of methylation at varying methyltransferase concentrations is again comparable between mutant and wild-type for hemi-methylated DNA, while unmethylated DNA is methylated 10 times more efficiently. That the linear fit of the rates of methylation plotted against the methyltransferase concentration does not intersect with the origin indicates an inactive methyltransferase at concentrations below 100 - 200 nM.

Given the apparent problems with and limitations imposed onto the methyltransferase assay it might have been better to try to develop an assay not based on steady state kinetics but on transient kinetics. An ideal scenario would have been to try to immobilise the methyltransferase on a carrier and have a continuous stream of substrates which could easily be varied and analysed. However the *EcoKI* methyltransferase turned out to be a very slow enzyme, as can be deduced from the exceedingly low turnover numbers (fast enzymes can reach turnover numbers of  $10^7$ ), and has thus shown linear progress curves which could easily be analysed. However it should be kept in mind that the enzyme might work far faster under a different set of reaction conditions.



### 4.3. Assessment of the stability of the methyltransferase complex at low enzyme concentrations by HPLC

Since the methyltransferase seemed to be inactive at low enzyme concentrations, I proposed that this might be due to the dissociation of the trimeric complex into its subcomponents. Since no free HsdS subunit had been observed before, a reaction of the type



seemed to be the most likely. D.T.F. Dryden investigated this possibility by HPLC gel filtration of the wild-type methyltransferase at different enzyme concentrations, and clearly demonstrated that the wild-type methyltransferase is not stable at lower protein concentrations, dissociating as expected into the dimeric form  $\text{M}_1\text{S}_1$  and free HsdM (Dryden *et al.*, 1997).

To investigate whether the mutant methyltransferases L113Q and L134V show the same behaviour, they too were applied at increasing dilutions onto an HPLC gel filtration column. The elution volume was recorded and the apparent molecular weights determined by comparison with a calibration curve supplied by D.T.F. Dryden.

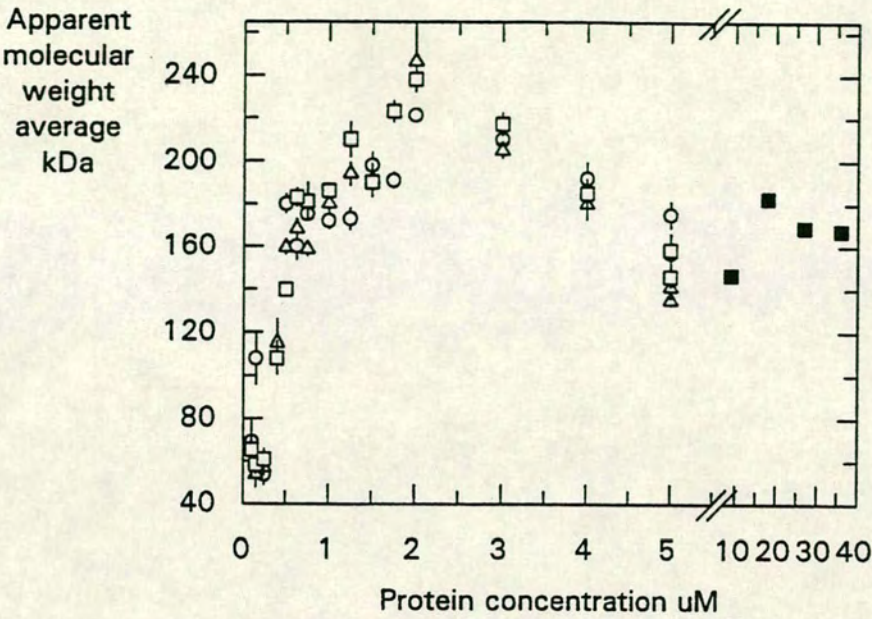
A plot of the apparent molecular weight against the enzyme concentration shows the decrease in the apparent molecular weight at protein concentrations below 300 nM, demonstrating the increasing instability of the mutant enzyme complexes as the enzyme concentration decreases.

This result explains why the methyltransferase showed no methylase activity at low enzyme concentrations, and validates the decision to increase the methyltransferase concentration in the standard methylation assay to 555 nM. It might also have important implications for the regulation of methylase activity *in vivo* as described in detail by Dryden *et al.* (1997).



**Fig. 4.5.1: HPLC Gel Filtration of wild-type Methyltransferase**

The apparent molecular weight of the EcoKI methyltransferase M<sub>2</sub>S<sub>1</sub> complex was measured by equilibrium sedimentation in an analytical ultracentrifuge.

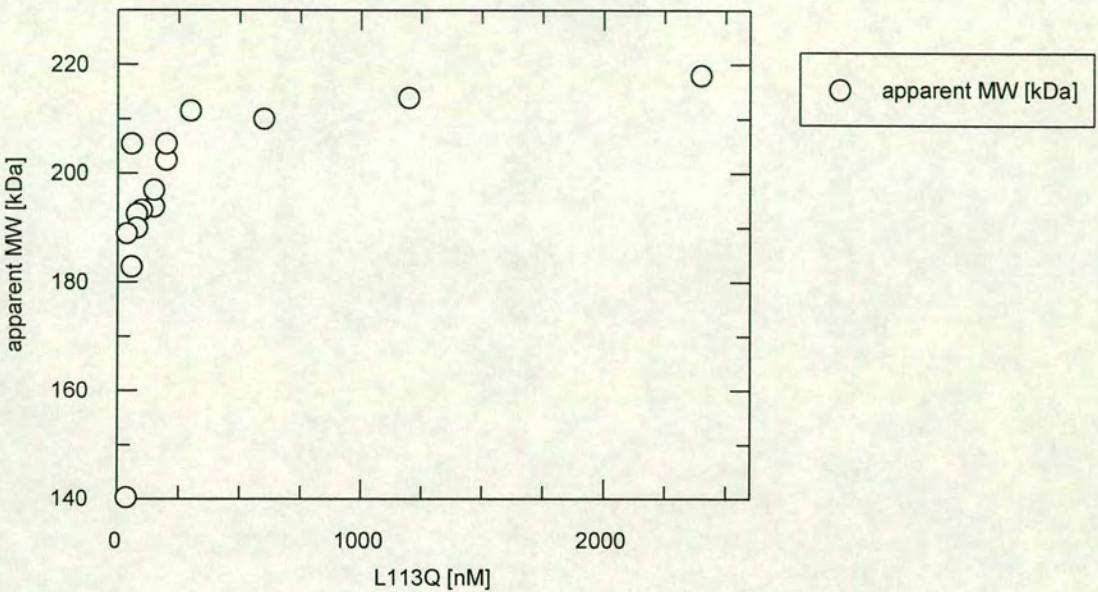


(Figure kindly provided by D.T.F. Dryden)



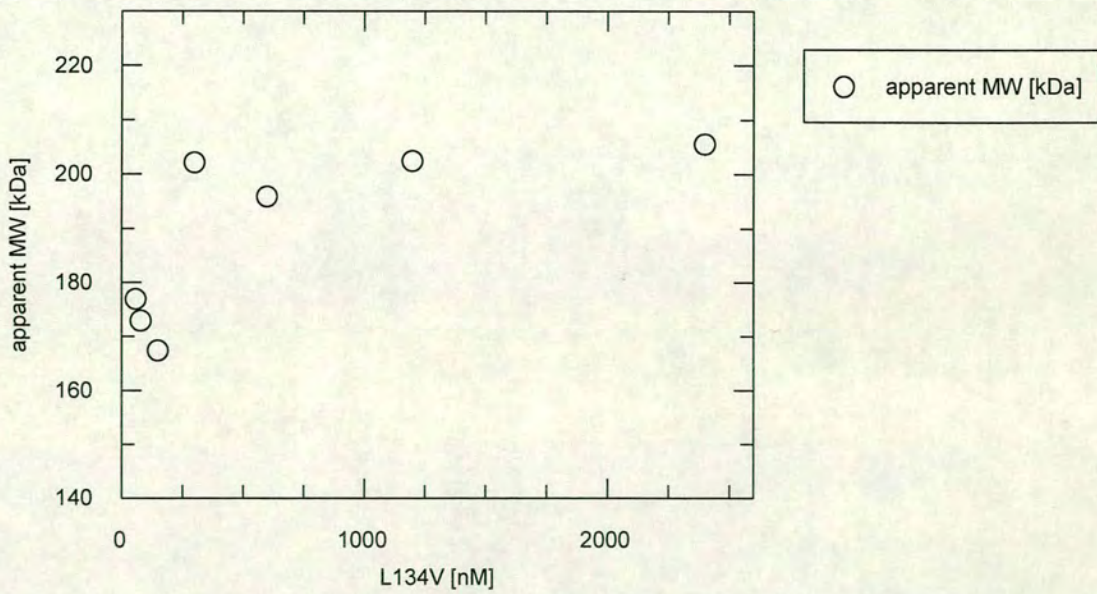
**Fig. 4.5.2: HPLC Gel Filtration of mutant Methyltransferase L113Q at increasing Dilutions**

The apparent molecular weight of the EcoKI methyltransferase  $M_2S_1$  complex was determined from the elution time on an HPLC gel filtration column.



**Fig. 4.5.3: HPLC Gel Filtration of mutant Methyltransferase L134V at increasing Dilutions**

The apparent molecular weight of the EcoKI methyltransferase  $M_2S_1$  complex was determined from the elution time on an HPLC gel filtration column.





## **CHAPTER 5:**

# **SEQUENCE ALIGNMENTS AND STRUCTURAL MODELLING**



## 5.1. Preface

The knowledge that C5-methyltransferases share ten well-defined motifs (Posfai *et al.*, 1988; Posfai *et al.*, 1989) was already established at the start of this work, and open reading frames from newly sequenced genes are accordingly classified as putative C5-methyltransferases even in the absence of experimental evidence if they contain these motifs. Open reading frames which do not contain these motifs are not regarded as potential C5-methyltransferases.

The situation was quite different for A6- and C4-methyltransferases, where alignments could only establish two conserved motifs (Klimasauskas *et al.*, 1989). Although more conserved regions were noted between highly similar sequences (see for example: Klimasauskas *et al.*, 1990; Bocklage *et al.*, 1991; Brooks *et al.*, 1991) and often implicated in sequence specific recognition (Kita *et al.*, 1992; Janulaitis *et al.*, 1992b), no comprehensive comparisons between all the available sequences were performed, nor were Type I or Type III systems included in the alignments. Since it seemed quite improbable that A6- and C4-methyltransferases should only contain two conserved motifs, and more likely that several subgroups of A6- and C4-methyltransferases were represented among the sequences, thereby hampering their alignment, it seemed possible that a detailed analysis of all the available sequences would establish additional motifs.

It was also hoped that these alignments would clarify how the different classes of methyltransferases could be brought together, and possibly shed some light on the relationship between Type I, Type II, and Type III enzymes, thereby substantiating how relevant information from research on Type II enzymes would be for research on other systems. This became especially important when the first crystal structure of a Type II C5-methyltransferase was published (Cheng *et al.*, 1993b).

Very little structural information was available for the HsdM subunit of Type I systems. Research in our own group on the *EcoKI* methyltransferase (Cooper and Dryden, 1994) had led to a proposed two domain model for HsdM (see Fig. 5.10) in which the N-terminal domain containing the m\* region and motif I is joined via a short flexible loop containing motif II to a C-terminal domain involved in subunit-subunit contacts with HsdS. Although the clustering of m\* mutations in the amino-terminal third of the HsdM subunit of *EcoKI* (Kelleher *et al.*, 1991) was not enough evidence



to suggest a separate domain for this region, it was hoped that the alignments would entail some structural information as well, thereby allowing further conclusions about the organisation of the HsdM subunit to be drawn

## **5.2. Sequence Alignments**

### **5.2.1. Limitations of alignment algorithms**

Sequence alignments have been done nearly exclusively using computers since the introduction of information technology into the field of molecular biology. The output produced - including a quote of the similarity and/or identity between the sequences as well as the probability of such a match - combined with the reputation of the computer as an objective calculator lends an air of measurable accurateness to the result. Conversely alignments done by hand are often dismissed as subjective and therefore likely to contain inaccuracies.

However the perceived accuracy of the computer-generated alignments tends to disguise the fact that alignments often are ambiguous (and can easily be modified by using different parameter settings), and that the alignments produced can only be as good as the algorithms applied to a specific problem, in this case a given set of sequences. Therefore it is important to be aware of the limitations of the program algorithms used. Equally important - but often neglected - is the question of how a given selection of sequences submitted for alignment affects the alignment itself.

The traditional and most commonly used method to align protein sequences with each other is to calculate the best overall alignment between them. This is usually done by a scoring system which awards points for similar or identical residues at a given position and incurs a penalty for mismatches and gaps in the alignment. The alignment with the highest score is considered the best match between the sequences. However this method can produce wildly inaccurate results even among closely related sequences if conserved regions are separated by spacers of variable length or if the conserved region comprises only a small part of the overall sequence (*e.g.* a shared domain with additional N- and/or C-terminal sequences).



More recent alignment methods have overcome this problem by searching for short conserved regions or motifs, usually by counting how often each combination of amino acids is found in a given set of sequences and checking for an above average occurrence. However this method is critically dependent on the right selection of sequences and will give rise to wrong results if unrelated sequences are included in the set.

Modern alignment methods tend to combine these two methods with each other. In a first step a database is searched for similar sequences which can be easily aligned with each other. From these alignments motifs are derived which in turn are used to find additional related sequences which were missed in the first search of the database. However while this combination neatly side-steps the problems associated with the respective algorithms by limiting itself to sets of fairly similar sequences (*e.g.* families and super families of sequences present in the database), it does not overcome their respective limitations.

Therefore while computer programs excel at the alignment of fairly similar sequences - in any case a tedious task - the best solution to the problem of how to align a set of protein sequences which show a common function but are not obviously related with each other (*e.g.* DNA methyltransferases) still requires the versatility and decision making capability of human input in at least both the selection process (*e.g.* "*Should this sequence be part of the alignment? Does it belong to a different group?*") and in the critical evaluation of the output produced (*e.g.* "*Is the alignment in this region relevant?*"), and is beyond the scope of the simple algorithms used today.

### **5.2.2. Preparation of sequences**

The EMBL, PIR, and GenBank sequence databases and their updates were routinely searched for entries containing either the words "methylase" or "methyltransferase" using the "stringsearch" program which is part of the UWGCG sequence analysis package. This method was superseded when the EMBL database went on-line and could be searched using a search form on the World Wide Web (at <http://www.ebi.ac.uk/srs/srsc>). In either case sequences were retrieved, imported into GeneJockey II™, the relevant open reading frames identified and translated (with due consideration of aberrant codon usages, *e.g.* "TGA" codes for the amino acid tryptophan in *Mycoplasma*), and the derived protein sequences examined for the



presence of motif I (FxGxG) and motif II (D/N/S P P F/Y) commonly found in A6- and C4-methyltransferases. Identified motifs were marked, and the sequences grouped according to their classification as  $\alpha$ -,  $\beta$ -, or  $\gamma$ -class enzymes (based on the relative position of the motifs to each other). The large number of RNA methyltransferase sequences retrieved were also examined and grouped accordingly.

### **5.2.3. Alignments by Computer**

Each of the methyltransferase sub-groups  $\alpha$ ,  $\beta$ , and  $\gamma$  were subjected to a variety of alignment methods, among them FASTA in the UWGCG package, Clustal in GeneJockey II<sup>TM</sup>, the Lipman-Pearson method in Lasergene DNA\*, and a weighted dynamic programming method called MaxHom (Sander and Schneider, 1991) as part of the PredictProtein/PHD program (Rost and Sander, 1993; Rost and Sander, 1994). However none of them produced satisfactory results despite several attempts with various parameter settings (*e.g.* gap penalty, window size, weight matrix), as could easily be judged by the failure of the programs to align the known conserved motifs.

Since these programs base their alignment on the best overall alignment score, the huge variability in both the spacing between the two motifs and the size of N- and C-terminal sequence regions seemed to be the likely reason for this failure, and it was decided to try to align the sequences by hand. It should be noted that motif-based alignment algorithms common today were still being developed and not available at the time.

### **5.2.4. Alignments by Hand**

The derived protein sequences were imported into the graphics program MacDraw Pro and grouped according to their classification as  $\alpha$ -,  $\beta$ -, or  $\gamma$ -class enzymes. The sequences were sorted within each class in ascending order by the number of amino acids present between the two known motifs and aligned at the first motif. Highly similar sequences were kept together, and obvious insertions noted as likely loop regions. Conserved residues were identified by eye, coloured accordingly, and aligned with each other. Alignments were continually updated when new



sequences became available, however the original alignments proved robust and did not need to be modified to accommodate the new sequences.

It became immediately obvious from these alignments that the *EcoRI* methyltransferase, which was supposed to belong to the  $\gamma$ -class, did not share any of the identified motifs with the other  $\gamma$ -class methyltransferases other than the ubiquitous motifs I and II (see Fig. 1d), and I therefore proposed it to be the first member of a new class.

#### 5.2.4.1. $\gamma$ -class

The alignment of all  $\gamma$ -class sequences (see Fig. 5.1) revealed ten conserved motifs, eight of which contain highly conserved and in some cases invariant residues. It also showed that both Type I and Type II methyltransferases of this class possess the same ten conserved motifs in the same order in their primary sequence, indicating greater similarities between these two types of methyltransferases than had previously been suspected. While the conserved motifs are normally located within the N-terminal half in Type II enzymes, they occupy the middle third of Type I HsdM polypeptides, and both the location and the size of the region of similarity, which at  $\approx 150$ -200 amino acids is comparable in size to domains in other proteins, are consistent with the view of Type I enzymes as modular systems that share a common methylation domain with Type II enzymes.

The variability in the number of residues located between the identified motifs was seen to indicate the location of loop regions, since additional residues could easily be accommodated at the surface of a common core structure. Similarly, the conservation of stretches of hydrophobic residues (*e.g.* in motif  $\gamma$ -VII and  $\gamma$ -X) was thought to indicate the presence of internal structures (probably  $\beta$ -sheet secondary structures), while the conservation of hydrophobic residues every third or fourth position (*e.g.* in motif  $\gamma$ -I and  $\gamma$ -II) would agree with an  $\alpha$ -helical structure located at the surface. And while most of the highly conserved residues lie at defined positions within a motif, two of the highly conserved amino acids (a leucine at the start of motif  $\gamma$ -VII and a phenylalanine between motifs  $\gamma$ -IX and  $\gamma$ -X) show some variability in their position, indicating their possible location in loop regions.

Type I methyltransferases share an additional highly conserved region between motifs  $\gamma$ -III and  $\gamma$ -IV. While this region is only present in two other  $\gamma$ -class methyltransferases, namely *MamI* and *BcgIA*, the highly conserved asparagine is also



found at the same position in C5-methyltransferases as part of the less conserved motif C5-II (Posfai *et al.*, 1989).

A comparison of all the identified motifs with the known conserved motifs of C5-methyltransferases revealed further similarities (see table 5.1), which implied that the conserved residues common to both might be involved with the same function in both types of methyltransferases, namely the binding of AdoMet. It further suggested that A6-methyltransferases of the  $\gamma$ -class are more similar to C5-methyltransferases than previously thought.

**Table 5.1: Comparison between some of the motifs identified in C5- and A6 $\gamma$ -methyltransferases**

Similar motifs occur in the same order in both C5- and A6 $\gamma$ -methyltransferases.

C5		A6 $\gamma$	
conserved residue	motif	motif	conserved residue
<b>xxD/S</b> <b>L</b> <b>F</b> <b>x</b> <b>G</b> <b>x</b> <b>G</b> <b>xx</b>	C5-I	$\gamma$ -II	<b>xxE/D</b> <b>P</b> <b>xx</b> <b>G</b> <b>x</b> <b>G</b>
N/S <b>E/D</b> <b>x</b> <b>d</b> <b>xx</b> <b>A</b> <b>xx</b> S/T <b>Y</b> <b>xx</b> N	C5-II	$\gamma$ -III	<b>G</b> <b>x</b> <b>E/D</b> <b>x</b> <b>d</b>
<b>D</b> <b>I/V</b>	C5-III	$\gamma$ -IV	<b>xxxxD</b> <b>f</b> <b>x</b>
<b>D/N</b> <b>xxxx</b> <b>G</b> <b>x</b> <b>P</b> <b>q</b> <b>xx</b> S <b>xx</b> G	C5-IV	$\gamma$ -V	<b>D</b> <b>xxxx</b> <b>N</b> <b>P</b> <b>P</b> <b>Y</b>
<b>D</b> <b>x</b> <b>r</b> <b>G</b> <b>x</b> <b>l</b> <b>xx</b>	C5-V	$\gamma$ -VI	<b>L</b> <b>x</b> <sub>3-4</sub> <b>G</b> <b>xxxxxx</b> <b>P</b> <b>xxxx</b>

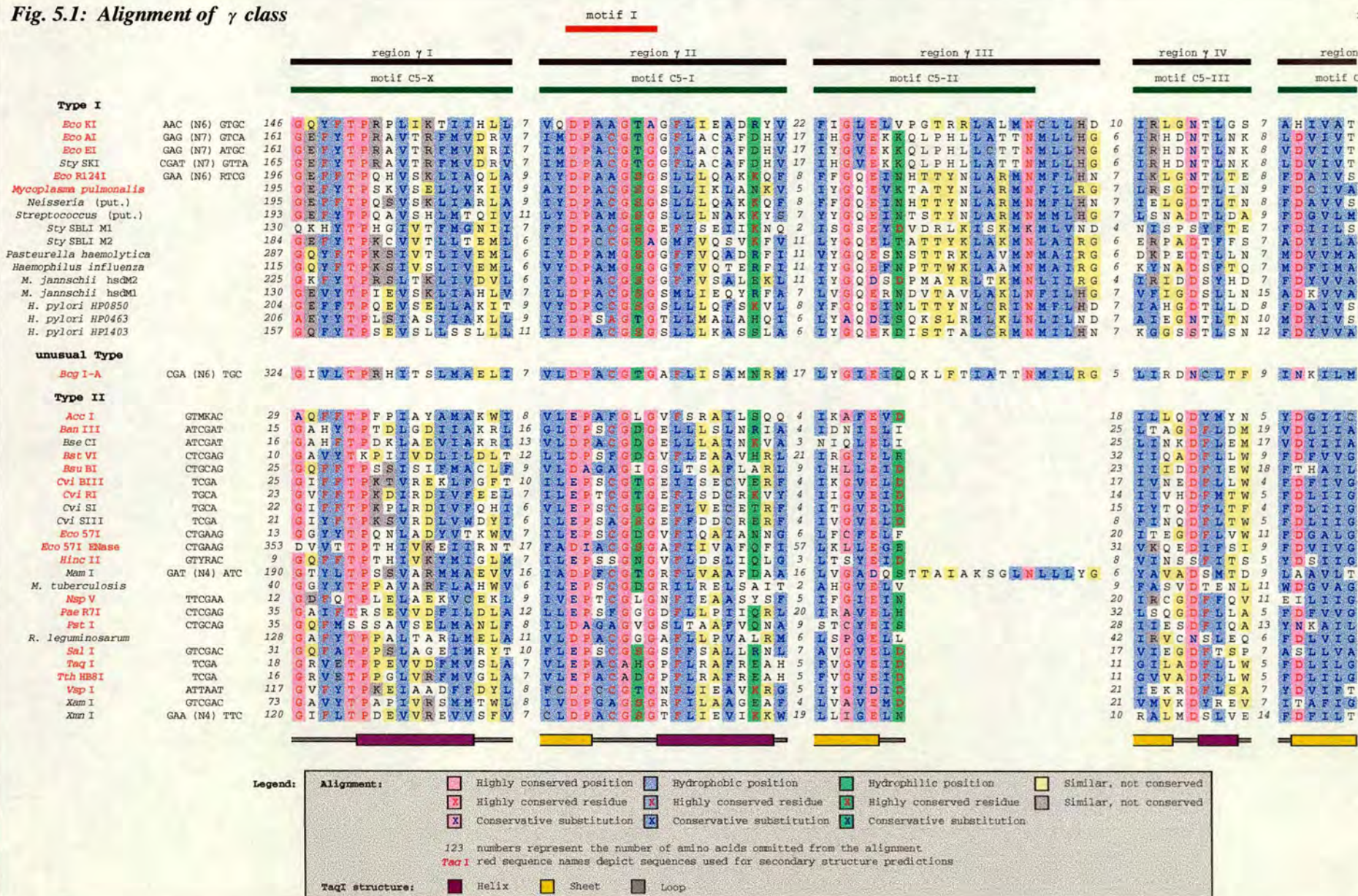
red: highly conserved residue - blue: hydrophobe residue  
lower case: preference for this residue - bold: similar between C5 and A6- $\gamma$

One major difference between C5- and A6 $\gamma$ -methyltransferases is the lack of a conserved region near the C-terminus in the A6 $\gamma$ -class. Instead an additional highly conserved region is located near the N-terminus in the A6 $\gamma$ -class, and was expected to perform a similar function to motif C5-X. As a consequence the linear order of motifs in the A6 $\gamma$ -class would be a circular permutation of the motif order in the C5-class.

The elucidation of the *TaqI* methyltransferase structure (Labahn *et al.*, 1994) and its comparison with the previously published catechol-O- (Vidgren *et al.*, 1994) and *HhaI* methyltransferase structures (Cheng *et al.*, 1993b) confirmed both the importance



Fig. 5.1: Alignment of  $\gamma$  class





\_\_\_\_\_

NPPYS<sup>9</sup> LSEL<sup>1</sup>SLFIQQALEM L VVG<sup>1</sup>GK<sup>1</sup>LCAIVPQSTMVGK NRHDK<sup>1</sup>AR<sup>1</sup>KKOILKQHTLETVITLNK<sup>1</sup>D T<sup>1</sup>THG VGVNPCIVIF KAGIKH<sup>110</sup>



of the identified conserved motifs and the conclusions drawn from their arrangement in the alignment. It also led to the proposal of a universal catalytic domain structure for AdoMet-dependent methyltransferases (Schluckebier *et al.*, 1995b), thereby lending support to the proposed common methylase domain for Type I and Type II methyltransferases.

#### 5.2.4.2. $\beta$ -class

The  $\beta$ -class of methyltransferases not only encompasses both Type II and Type III systems, but includes A6- and C4-methyltransferases. Despite this heterogeneity their alignment (see Fig. 5.2) revealed a remarkably homogeneous group with seven regions of similarity, some of them quite extended and containing more than one set of conserved residues.

As in the  $\gamma$ -class of A6-methyltransferases conserved patterns of hydrophobic and hydrophilic residues emerged in the alignment that strongly indicate the location of  $\beta$ -sheet and  $\alpha$ -helical structures respectively, while regions of variable length between the conserved regions are again indicative of loop regions.

Contrary to the  $\gamma$ -class enzymes (with the exception of *MamI* and *VspI*), where the region of homology lies near the N-terminus and is followed by a large variable region at the C-terminus which is believed to be the target recognition domain, the variable region in the  $\beta$ -class enzymes is less easily pinpointed. While some have extended N-terminal (*e.g.* *BamHI*, *CfrBI*, *Pac25I*) or C-terminal (*e.g.* *HpaI*, *HinfI*, *CcrI* and its homologues from *B. abortus* and *S. meliloti*) variable regions, others have neither (*e.g.* *BstF5I*, *Cfr9I*, *DpnII*, *EcoVIII*, *HindIII*). However in these cases an extended variable region is found between motif  $\beta$ -V and  $\beta$ -VI, sometimes in combination with another extended variable region after motif  $\beta$ -II or  $\beta$ -III. It is as yet unclear how the different  $\beta$ -class enzymes bind to their respective DNA target, but it seems certain that different enzymes use different regions (and probably different modes) to bind DNA.

A comparison of the conserved regions with the conserved regions of A6 $\gamma$ -methyltransferases (see table 5.2) revealed similar motifs. As in the comparison of the A6 $\gamma$ - with the C5-methyltransferases not all regions could be assigned to equivalent regions in the other methyltransferase classes. However the arrangement of the motifs included in the comparison again exhibits a circular permutation, with motif  $\beta$ -III (equivalent to C5-III) closest to the N-terminus. Since the assignment is



Fig. 5.2: Alignment of  $\beta$  class



Legend:

Alignment:



Highly conserved position



Hydrophobic position



Hydrophilic position



Similar, not conserved



Highly conserved residue



Highly conserved residue



Highly conserved residue



Similar, not conserved



Conservative substitution



Conservative substitution



Conservative substitution



Similar, not conserved



motif I

region  $\beta$  V

region  $\beta$  VI

region  $\beta$  VII

motif C5-VIII

motif C5-X

motif C5-I

motif C5-II

NH 4 FSDTH EY IYVYA K 157 FEGPK PVP LITD LVKIGTK KDSLV L DFFAGS GTTAE AVAYLNEK DSGCRNFIC IQKDEVINK 176  
 NP 2 IAVEHE YIIVYT K 165 FTNAKT IKLVEDL ISFACD GEGIV L DFFAGS GTTAE HTVFNLNNEK NKT SYQFITV QLDDEPK K 169  
 SN 4 VSIQGE YILVYA K 129 FDT PK P TALLKKI IKLAID KDGVL L DFFAGS GTTAE HAVMALNEE DGGQRTFILCT IDQALS N 200  
 VS 7 LKDS T EFINVFS K 181 KNGK K P EALLRR I I DMTTK EGDIV L DFFAGS GTTAE AVAHKMNEC YIGIE QMDYI ET LAVER R 136

IR 24 ARSCW E P WGMFR K 45 HPSLKP QRFLRQ VVRAALPL GIG I I Y DFFAGS GSTLA AAEAV GYFAIQ TDRDAQYFGIGTKAF 10  
 PI 3 LYPANY SLLYYV K 59 KFNELE SVKLLDR IITMSTN EGDIV L DFFAGS GTTAE AVSEM LGRKKYIGFELGNCEI I KERLKN 63  
 HA 3 FAGRH E TILWFT K 58 HPTQY P P EDMIE I VLSSTTE PNDIV L DFFAGS GTTAE ASVAKNLNRYFYQAE IEKE YVDIAYQIL 78  
 PR 5 YIPNV EMIQWYV K 31 HPTQK P PVKLIBEY LIRIHSN ENDI I L DFFAGS GTTAE GVASLNLNRYFYQAE IEKE YVDIAYQIL 8  
 PN 5 FANAHE TLIWAS K 46 HPTQK P P EALLRYR VILSTTK PGDVI L DFFAGS GTTAE GAAAKRLGRKKYIGFELGNCEI I KERLKN 108  
 PN 5 FQNAHE TLIWAS R 46 HPTQK P P EALLARILMMAS K PGDVI L DFFAGS GTTAE GAVAKRLGRKKYIGFELGNCEI I KERLKN 112  
 AE 5 FQNAHE TLIWATA 46 HPTQK P P EALLARILMMAS K PGDVI L DFFAGS GTTAE GAVAKRLGRKKYIGFELGNCEI I KERLKN 111  
 PE 5 PTRSH E FLFMLT K 37 HFAT F P TELIRP C I LASTK PGDVI L DFFAGS GTTAE GAVAKRLGRKKYIGFELGNCEI I KERLKN 13  
 GW 18 PVPVT EYILVYR K 43 HPAT F P YGLAERVIKYYSF KNDVI L DFFAGS GTTAE AKAAIDLGRRFVMC EISKQYIDLII EGT 10  
 PN 5 FTHST E TILWAS K 41 HPTQK P P EYLLER I I LASTK EGDY I L DFFAGS GTTAE GVVAKRLGRKKYIGFELGNCEI I KERLKN 7  
 NS 4 FANLNDY ILCYS K 118 YPTEKN FNMMLK L IVGASSN PGDLV I L DFFAGS GTTAE LHAASLLQRRKWIGFELGNCEI I KERLKN 69  
 HR 27 LQPT F EPILVFS K 47 HPTQK P P VALMKKT L I ELTTQ KGQLV I DFFAGS GTTAE LVAAKDLGRDYIGFELGNCEI I KERLKN 2  
 MG 2 TRRRS EYLVV I Q K 26 HTHSK P P IEMQKQ L I ELTTQ EGDY I L DFFAGS GTTAE GVSFECCKQTNRNFI GCDLI FGDDENEQD 14  
 HR 27 LRPL F EPILVFS K 52 HVAQK P P LNLMLK L I ELTTQ EEQIV L DFFAGS GTTAE LLAAKELNRHFI GY EKNNGIYNIAVNR 14  
 PN 5 FCNAHE TMLWCS K 45 HSTQK P P ESLLYK V I LSSSK PNDIV L DFFAGS GTTAE GAVAKALGRKKYIGFELGNCEI I KERLKN 105  
 GK 3 YSPRH DDILMFT K 49 HPTQK P P EALYERMI LASSN EGDIV L DFFAGS GTTAE LNFVCKHLNRSGIGIDINK EYIEMAKER 80  
 KG 4 IASAH EYLLVYG K 144 FDT PK P KALNYIMS I INCMAK PDALI L DFFAGS GTTAE HAAAVLNSLDGGS RKTILMESNHPITK 38  
 PN 5 FTHST E TILWAS K 41 HPTQK P P EYILER I I LASTK ENDI I L DFFAGS GTTAE GVVAKRLGRKKYIGFELGNCEI I KERLKN 8  
 PN 5 PTYSS EWI IWAR K 40 HPTQK P P LGLLSRI I LSSSTQ KDDL I L DFFAGS GTTAE GIAGVLLDRNFI GY EKNNGIYNIAVNR 22  
 GS 4 FSTGQ E TILFFS K 68 HITPK P P RDLIER I I RASSN PNDIV L DFFAGS GTTAE I VAKKLGRKKYIGFELGNCEI I KERLKN 6  
 PN 13 VTLDV EYILIFR K 45 RTASF P P EIPRR L I RMFS I IGDIV L DFFAGS GTTAE VKAAIE LKRNFI GY EKNNGIYNIAVNR 19  
 QK 26 LSIMH EYIFHRT K 52 HPAT F P P KLP E M C I KLHG V K KTNL V L DFFAGS GTTAE STAIACIRLGI DYIGFELGNCEI I KERLKN 26  
 GF 43 AQPPH LGAIGA H K 25 NETQK P P VALVEP LVRNACP PGGLV A DFFAGS GTTAE STALACLATGRRFI GY EKNNGIYNIAVNR 10  
 PT 13 VTHHE EYILIFR K 50 RAAAY F P FELAYR L INMYSI MG DVI L DFFAGS GTTAE IAAACAGRNFI GY EKNNGIYNIAVNR 81  
 PE 5 PARSH EFLFMLT K 37 HFAT F P PDLIRP C I LASTK PGDVI L DFFAGS GTTAE VGLVVCQDENRQYIGFELGNCEI I KERLKN 13  
 FP 4 YYQIF EYMFIFS K 70 HPAIF P P EKLAE DHI LSWSN EGDIV L DFFAGS GTTAE AKMAALNRRKKYIGFELGNCEI I KERLKN 22  
 PS 11 VKDRA VNTVWFS K 80 HPAR F P P AKLP E FFI LMLTE PDDL V L DFFAGS GTTAE VGLVVCQDENRQYIGFELGNCEI I KERLKN 29  
 SA 3 PANRH E IAWFA K 58 HPTQK P P AAVIER LVRALSH PGSTV L DFFAGS GTTAE VTARVA IQEGRNFI GY EKNNGIYNIAVNR 35  
 PE 5 PTRSH EYIFLFS K 39 HFAV F P P R A M A R L C V L A G S R PGKVL L DFFAGS GTTAE GVVCQDELRECVIGFELGNCEI I KERLKN 4  
 PE 5 PARSH EFLFME T K 37 HFTT F P P E L I R P C I H A S T E PGDVI L DFFAGS GTTAE VGLVVCQDENRQYIGFELGNCEI I KERLKN 13



continuous and not interrupted by other conserved regions, it is tempting to speculate that the other motifs will follow in linear order, but the alignment itself neither supports nor contradicts this speculation.

**Table 5.2: Comparison between some of the motifs identified in the  $\beta$ -class with C5- and A6 $\gamma$ -methyltransferases**

Similar motifs occur in permuted order in C5-, A6 $\gamma$ -, and  $\beta$ -class methyltransferases.

A6 $\gamma$		C5	$\beta$ -class	
conserved residue	motif	motif	motif	conserved residue
<b>x</b> x x x <b>D</b> f <b>x</b>	$\gamma$ -IV	C5-III	$\beta$ -I	<b>x</b> x x x <b>D/N</b> x x x x x x x
<b>D</b> x x x x <b>N P P Y</b>	$\gamma$ -V	C5-IV	$\beta$ -II	x <b>D</b> x x x x <b>D/S P P Y</b>
<b>L</b> x <sub>3-4</sub> <b>G</b> x x x x x x <b>P</b> x x x x	$\gamma$ -VI	C5-V	$\beta$ -III	<b>L</b> x <sub>3</sub> <b>G</b> x x x x x x x
<b>G</b> x <b>F/Y F/Y T P</b>	$\gamma$ -I	C5-X ?	$\beta$ -VI	<b>H</b> p t q x <b>P</b>
<b>x</b> x <b>E/D</b> <b>P</b> x x <b>G</b> x <b>G</b>	$\gamma$ -II	C5-I	$\gamma$ -VII	<b>D</b> x x <b>l D P F</b> x <b>G S G T T</b>
<b>G</b> x <b>E/D</b> x <b>d</b>	$\gamma$ -III	C5-II	$\beta$ -VII	<b>R</b> x x i <b>G</b> x <b>E/D</b> i <b>d/n</b>

red: highly conserved residue - blue: hydrophobe residue  
lower case: preference for this residue - bold: similar between A6 $\gamma$  and  $\beta$ -class

The recent publication of the *PvuII* methyltransferase structure (Gong *et al.*, 1997) not only confirmed the importance of all the identified conserved regions (except region  $\beta$ -IV), but also their assignment in regard to A6 $\gamma$ - and C5-methyltransferases. That the high conservation of region  $\beta$ -IV was only recognised for A6-, but not for C4-methyltransferases is probably the reason why the equivalent to motif C5-VI was not located in the structure.

The similarity of the methylation domain of M.*PvuII* with the methylation domains in previously published methyltransferase structures lends strong support to the idea of an universal catalytic domain structure for AdoMet-dependent methyltransferases. Even so an answer to the problem of how different polypeptide chains - which sometimes harbour large insertions in different places - can fold into the same structure is not apparent.



### 5.2.4.3. $\alpha$ -class

Although the  $\alpha$ -class was expected to be the most homogeneous group since it only contains simple monomeric A6-methyltransferases (albeit both Type II enzymes and *dam* homologues), it proved the hardest to align (see Fig. 5.3). This was only in part due to the fact that the  $\alpha$ -class is the smallest group and accordingly had the fewest protein sequences available for alignment. More significant was the fact that the  $\alpha$ -class actually consists of two sub-classes which share extensive similarities with each other near the N-terminus, but little similarity in the rest of the sequence. Aggravating, too, was the fact that most sequences in the sub-class  $\alpha$ -1 are homologues of or recognise the same target sequence as the *dam* methylase, making it difficult to distinguish conserved regions involved in methylation from similarly conserved regions involved in DNA binding.

A total of twelve conserved regions was identified in both subclasses. As in the other methyltransferase classes conserved patterns of hydrophobic and hydrophilic

**Table 5.3: Comparison between some of the motifs identified in the  $\alpha$ -classes with C5-methyltransferases**

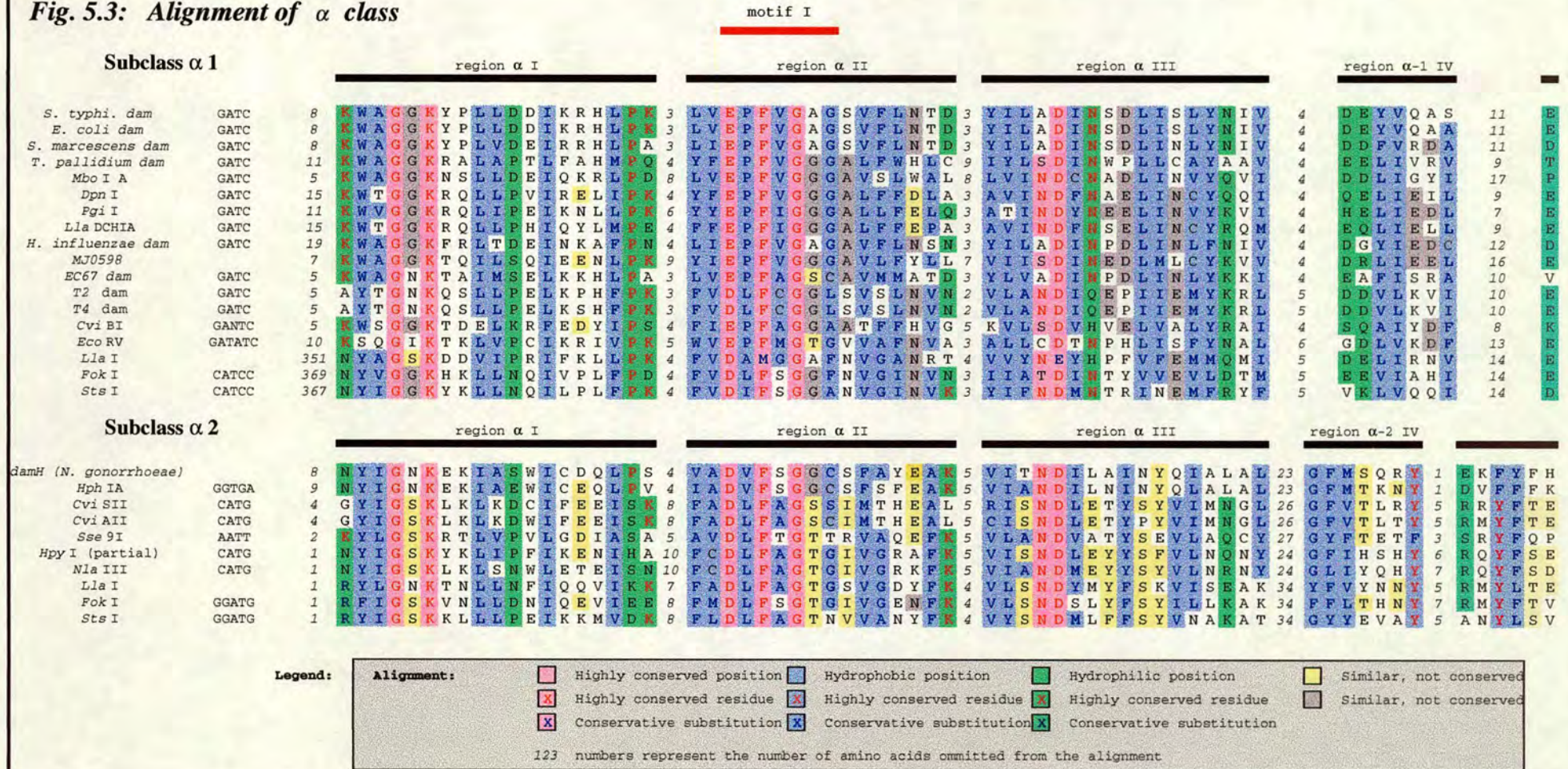
Similar motifs occur in C5- and  $\alpha$ -class methyltransferases.

A6 $\alpha$ -1		C5	A6 $\alpha$ -2	
conserved residue	motif	motif	motif	conserved residue
k <b>xxGgK</b> xxxxxxxxxxx	$\alpha$ -I	C5-X ?	$\alpha$ -I	xxx <b>GsK</b> xxxxxxxxxxx
xx <b>E/DpF</b> xx <b>gg</b> axxxxx	$\alpha$ -II	C5-I	$\alpha$ -II	xx <b>DlF</b> xx <b>Gt</b> gxxxxxxxx
xxx <b>nD</b> xxxxxxxxxxxxx	$\alpha$ -III	C5-II ?	$\alpha$ -III	xxx <b>ND</b> xxxxxxxxxxx
xxxd <b>F</b> x	$\alpha$ -1 VII	C5-III ?	$\alpha$ -2 VII	xxx <b>D</b> ax
d <b>xxxxDPPY</b>	$\alpha$ -VIII	C5-IV	$\alpha$ -VIII	d <b>xxxxDPPY</b> n
<b>E</b> xxxxxx	$\alpha$ -XII	C5-II ?	$\alpha$ -XII	<b>E</b> xxxxxx

red: highly conserved residue - blue: hydrophobe residue  
lower case: preference for this residue - bold: similar between A6 $\alpha$ -1 and A6 $\alpha$ -2



Fig. 5.3: Alignment of  $\alpha$  class





motif II

region α-1 V										region α-1 VI										region α-1 VII										region α VIII																																																			
V	Y	Y	Q	L	R	E	E	F	N	11	F	L	Y	L	N	R	Y	G	Y	N	G	L	C	R	Y	N	L	R	G	E	F	N	V	P	F	G	R	Y	K	R	P	Y	F	P	E	A	E	L	2	F	A	E	K	A	Q	N	A	F	P	Y	C	E	S	Y	A	5	A	D	K	S	S	V	V	Y	C	D	P	P	Y	A	
V	Y	Y	Q	L	R	E	E	F	N	11	F	L	Y	L	N	R	Y	G	Y	N	G	L	C	R	Y	N	L	R	G	E	F	N	V	P	F	G	R	Y	K	K	P	Y	F	P	E	A	E	L	2	F	A	E	K	A	Q	N	A	F	P	Y	C	E	S	Y	A	5	A	D	D	A	S	V	V	Y	C	D	P	P	Y	A	
Q	Y	R	L	A	R	E	E	F	N	11	F	L	Y	L	N	R	H	C	Y	N	G	L	C	R	Y	N	L	R	G	E	F	N	V	P	F	G	R	Y	K	K	P	Y	F	P	E	E	E	L	2	F	A	E	K	S	R	N	A	T	F	V	C	E	H	Y	R	5	A	A	A	G	A	V	V	Y	C	D	P	P	Y	A	
Y	Y	R	L	A	R	R	K	F	A	11	F	L	Y	L	N	R	S	C	Y	N	G	L	C	R	Y	N	K	A	G	Q	F	N	V	P	F	L	G	R	A	A	P	A	S	P	F	L	N	T	T	27	C	A	R	A	L	A	N	T	T	L	N	C	Q	H	F	S	3	F	A	R	G	D	F	V	Y	L	D	P	P	Y	L
Y	F	Y	H	K	R	D	V	F	N	12	F	I	F	L	N	K	S	A	F	N	G	L	C	R	Y	N	K	N	N	Q	F	N	V	P	I	G	N	Y	K	K	P	T	F	V	D	K	E	N	3	I	S	K	K	L	Q	N	T	K	I	L	S	G	D	F	E	7	N	N	F	P	C	L	F	Y	L	D	P	P	Y	R	
Y	Y	L	D	L	R	S	A	D	R	15	I	L	Y	M	L	R	V	N	F	N	G	L	C	R	Y	N	S	K	N	Q	F	N	V	P	Y	G	R	Y	K	N	P	K	I	V	D	E	E	L	5	V	Y	I	N	N	N	Q	L	E	I	K	V	G	D	F	E	5	V	R	T	G	D	F	V	Y	F	D	P	P	Y	I	
Y	F	Y	E	I	R	A	I	D	R	15	I	I	Y	L	N	K	T	C	Y	N	G	L	C	R	Y	N	N	A	G	E	F	N	S	P	F	G	K	Y	K	N	P	N	I	V	N	E	P	V	5	K	Y	L	N	T	A	K	I	Q	I	F	N	G	D	Y	Q	5	I	P	R	S	S	F	V	Y	L	D	P	P	Y	H	
Y	Y	L	D	L	R	S	S	D	R	15	I	M	Y	M	L	R	V	D	F	N	G	L	C	R	Y	N	S	K	N	Q	F	N	V	P	Y	G	R	Y	K	N	P	K	I	V	D	K	E	L	5	E	Y	L	N	N	S	I	K	I	M	S	G	D	F	E	5	A	Q	D	G	D	F	V	Y	F	D	P	P	Y	I		
Y	Y	Y	A	K	R	R	Q	F	N	11	F	L	Y	L	N	R	F	G	F	N	G	L	C	R	Y	N	S	K	N	E	F	N	V	P	F	G	A	Y	K	T	H	Y	F	P	E	D	E	L	2	F	A	H	K	A	Q	S	A	V	F	L	C	C	D	F	Q	5	A	D	K	D	S	V	I	Y	C	D	P	P	Y	A	
F	Y	Y	K	V	R	D	D	F	N	14	F	I	F	L	N	K	T	C	Y	N	G	L	C	R	Y	N	K	K	G	E	F	N	V	P	Y	G	R	Y	K	N	P	K	I	F	D	E	O	N	3	V	S	K	L	L	K	N	V	K	I	L	C	G	D	F	E	5	V	D	A	E	S	F	V	Y	F	D	P	P	Y	K	
A	Y	Y	N	I	R	Q	E	F	N	13	F	L	Y	L	N	R	H	G	Y	R	G	L	C	R	Y	N	K	S	G	H	F	N	I	P	F	G	N	Y	K	N	P	Y	F	P	E	K	E	I	2	F	A	E	K	A	Q	R	A	T	F	I	C	A	S	F	D	5	L	Q	V	G	D	V	V	Y	C	D	P	P	Y	D	
E	F	L	K	L	R	E	D	Y	N	7	L	Y	V	L	H	F	H	G	F	S	N	M	I	R	I	N	D	K	G	N	F	T	T	P	F	G	K	R	T	I	N	K	N	S	E	K	R	F	2	F	K	Q	N	C	D	K	I	I	F	S	S	L	H	F	K	3	I	L	D	G	D	F	V	Y	V	D	P	P	Y	L	
E	F	L	K	L	R	E	D	Y	N	7	L	Y	V	L	H	F	H	G	F	S	N	M	I	R	I	N	D	K	G	N	F	T	T	P	F	G	K	R	T	I	N	K	N	S	E	K	R	F	2	F	K	Q	N	C	D	K	I	I	F	S	S	L	H	F	K	3	I	L	D	G	D	F	V	Y	V	D	P	P	Y	L	
T	Y	Y	E	V	R	S	W	K	P	9	F	Y	Y	L	R	K	T	C	F	R	G	M	M	R	Y	N	K	N	G	G	F	N	V	P	F	G	R	Y	K	T	Y	N	F	E	D	I	I	N	2	Y	Y	N	I	L	K	D	T	I	I	L	E	K	S	F	D	6	N	D	S	S	N	F	V	F	L	D	P	P	Y	D	
Y	Y	Y	E	V	R	E	R	F	N	7	F	L	F	L	N	R	S	C	F	N	G	M	I	R	F	N	S	K	G	G	F	N	V	P	F	C	K	K	P	N	R	F	A	Q	A	Y	I	T	10	E	I	I	S	K	G	N	Y	T	F	L	C	Q	S	F	E	5	V	N	R	D	D	V	V	Y	C	D	P	P	Y	I	
A	F	N	R	L	R	D	H	Y	N	8	L	Y	T	L	N	I	Y	S	F	Q	N	I	L	R	F	N	Q	A	K	K	Y	N	T	P	I	G	N	N	E	F	N	E	G	Y	K	D	R	I	7	P	E	V	E	M	R	L	G	S	Y	S	A	I	N	F	N	1	Y	D	D	D	T	V	F	Y	F	D	P	P	Y	L	
G	F	K	A	F	R	N	Y	Y	N	7	L	Y	T	L	I	C	Y	S	F	N	Y	Q	F	R	F	N	N	N	Q	E	Y	N	N	P	F	G	R	E	R	S	Q	F	S	P	A	L	K	K	11	K	N	V	Q	F	V	C	S	E	F	E	H	F	N	F	S	1	L	D	Q	N	D	L	V	Y	C	D	P	P	Y	L	
A	F	L	A	F	R	K	H	Y	N	7	L	Y	V	L	S	S	F	S	Y	N	Y	Q	F	R	F	N	S	M	E	F	N	N	P	F	G	R	N	R	S	H	F	S	E	N	M	R	N	11	L	D	A	T	F	T	D	N	Y	F	N	E	F	D	F	S	1	L	T	S	N	D	F	I	Y	L	D	P	P	Y	L		

region α-2 V										region α-2 VI										region α-2 VII										region α VIII																																																					
D	E	C	Q	Q	L	D	L	Y	R	2	V	S	T	D	E	Y	N	F	L	L	A	S	L	L	T	S	A	D	S	V	K	N	T	S	V	V	Y	G	A	Y	L	K	K	F	K	K	T	A	L	K	R	M	V	F	T	P	L	11	F	M	O	D	I	F	D	L	L	2	I	Q	A	D	A	V	Y	L	D	P	P	Y	T		
E	E	C	R	E	L	D	A	I	R	2	V	S	T	D	E	Y	N	F	L	L	A	S	L	L	T	S	A	D	S	V	K	N	T	S	V	V	Y	G	A	Y	L	K	K	F	K	K	T	A	L	K	R	M	V	F	T	P	L	11	Y	N	E	D	V	F	E	L	L	2	I	Q	A	D	A	V	Y	L	D	P	P	Y	T		
D	N	A	M	K	I	D	I	I	R	2	V	S	T	D	E	Y	N	F	L	L	A	S	L	L	T	S	A	D	S	V	K	N	T	S	V	V	Y	G	A	Y	L	K	K	F	K	K	T	A	L	K	R	M	V	F	T	P	L	11	F	M	E	D	A	T	E	L	G	2	I	K	T	D	I	A	Y	V	D	P	P	Y	N		
D	I	A	M	R	I	D	I	I	R	2	V	S	T	D	E	Y	N	F	L	L	A	S	L	L	T	S	A	D	S	V	K	N	T	S	V	V	Y	G	A	Y	L	K	K	F	K	K	T	A	L	K	R	M	V	F	T	P	L	11	F	M	E	D	A	T	E	L	G	2	I	K	T	D	I	A	Y	V	D	P	P	Y	N		
K	N	G	A	R	V	D	A	I	R	4	A	D	S	W	M	R	P	I	L	L	T	S	L	M	L	A	A	D	R	V	D	S	T	T	G	V	Q	M	A	Y	L	K	Q	W	A	A	R	A	H	N	D	L	E	L	R	L	P	9	A	R	E	D	A	V	T	L	A	3	P	R	V	Q	L	M	Y	L	D	P	P	Y	N		
T	N	A	Q	K	I	D	A	V	R	4	I	D	N	C	A	Y	Y	F	L	L	A	S	L	L	E	S	A	D	K	V	A	N	T	A	S	V	Y	G	A	F	L	K	N	L	K	K	S	A	L	K	P	L	S	L	E	P	A	11	Y	Q	Q	D	A	S	E	L	I	2	I	S	G	D	I	L	Y	L	D	P	P	Y	N		
E	N	G	K	K	I	D	A	V	R	4	I	D	E	D	T	Y	Y	F	L	L	A	T	L	L	E	G	A	D	K	V	A	N	T	A	S	V	Y	G	A	F	L	K	N	L	K	K	S	A	L	K	P	L	S	L	E	P	A	11	Y	Q	Q	D	A	N	Q	L	I	2	I	S	G	D	I	L	Y	L	D	P	P	Y	N		
E	N	A	L	K	I	D	G	M	R	4	I	S	K	A	E	Y	S	Y	L	L	A	S	L	L	E	S	A	D	V	T	K	V	S	N	T	S	G	T	Y	Q	A	F	F	K	F	W	E	S	R	A	L	K	K	F	T	I	M	P	L	14	F	M	K	N	T	N	R	L	V	2	I	S	G	D	I	A	Y	I	D	P	P	Y	T
E	N	A	S	R	I	D	F	I	R	4	I	N	E	L	E	F	A	Y	L	L	A	I	L	I	E	A	V	P	F	I	S	N	I	S	G	T	Y	Q	A	Y	L	K	H	W	D	K	R</																																				



region $\alpha$ -1 IX										region $\alpha$ -1 X										region $\alpha$ -1 XI										region $\alpha$ XII																																																	
PLS	A	T	A	N	F	T	A	Y	4	F	S	L	T	Q	Q	A	H	L	A	E	I	A	E	N	L	V	S	N	R	I	P	V	L	I	S	N	H	3	L	T	R	E	W	Y	Q	L	A	K	L	H	V	V	K	V	R	P	S	I	S	S	N	G	G	T	R	K	K	2	E	L	L	A	L	Y	Q	9			
PLS	A	T	A	N	F	T	A	Y	4	F	T	L	E	Q	Q	A	H	L	A	E	I	A	E	G	L	V	E	R	H	I	P	V	L	I	S	N	H	3	L	T	R	E	W	Y	Q	R	A	K	L	H	V	V	K	V	R	R	S	I	S	S	N	G	G	T	R	K	K	2	E	L	L	A	L	Y	K	9			
PLS	A	T	A	N	F	T	A	Y	4	F	S	I	A	E	Q	Q	A	H	L	A	E	I	A	E	G	L	V	E	R	S	R	V	P	V	L	I	S	N	H	3	L	T	R	D	W	Y	Q	H	A	A	L	Y	V	V	K	V	R	R	T	I	S	R	N	I	L	G	R	S	K	2	E	L	L	A	L	Y	R		
										F	D	R	A	A	H	E	S	L	A	A	F	C	M	H	L	D	A	R	G	V	L	F	M	L	S	N	S	3	E	V	R	A	W	Y	R	P	F	R	V	Q	Q	L	N	A	P	R	C	I	A	R	S	A	H	A	R	G	K	2	E	V	L	I	T	N	Y	8			
P	I	S	D	T	A	S	F	T	S	Y	4	F	D	D	N	E	Q	K	R	L	A	N	F	C	K	K	I	D	K	L	G	H	Y	F	L	L	S	N	S	10	F	F	D	E	L	Y	Q	D	F	K	I	E	R	I	Q	A	N	R	T	I	S	A	N	S	N	G	R	K	K	2	E	I	I	V	S	N	G	1	
PLS	E	T	S	A	F	T	S	Y	4	F	S	F	A	D	Q	V	R	L	R	D	A	F	K	R	L	S	D	T	G	A	Y	V	M	L	S	N	S	3	L	V	E	E	L	Y	K	D	F	N	I	H	Y	V	E	A	T	R	T	N	G	A	K	S	S	R	G	K	2	E	I	I	V	T	N	Y	2				
P	I	S	Q	S	A	N	F	T	G	Y	4	W	D	E	K	D	Q	I	R	L	R	N	V	C	N	T	L	N	E	R	G	I	K	F	L	L	S	N	S	3	F	I	K	E	I	Y	S	D	Y	N	I	Y	V	V	Q	A	T	R	A	V	N	S	D	S	S	K	R	G	Q	2	E	F	F	L	I	N	N	Y	1
PLS	E	T	S	A	F	T	S	Y	4	F	S	Y	E	D	Q	V	R	L	R	D	A	F	K	Q	L	D	S	K	G	V	F	V	M	L	S	N	S	3	L	A	E	E	I	Y	K	D	F	N	I	H	K	I	E	A	T	R	T	N	G	A	K	S	S	R	G	K	2	E	I	I	V	T	N	Y	2				
PL	Q	Q	E	T	N	F	T	G	Y	4	F	G	L	A	Q	Q	R	A	L	A	D	L	A	K	S	I	Q	K	E	K	Q	I	S	I	L	I	S	N	H	3	F	T	R	E	I	Y	N	G	A	K	F	K	R	V	K	V	Q	R	S	I	S	Q	N	P	E	K	R	V	K	2	E	L	I	A	I	F	G	3	
PL	N	K	T	S	F	T	S	Y	4	F	N	D	D	Q	I	R	L	A	K	F	Y	R	K	L	D	K	R	G	A	K	L	M	L	S	N	S	4	F	F	G	K	L	Y	E	G	F																																	



residues in the alignment strongly indicated the location of  $\beta$ -sheet and  $\alpha$ -helical structures respectively, while regions of variable length between the conserved regions were regarded as likely loop regions.

Among the five highly conserved regions common to both subclasses region  $\alpha$ -II and  $\alpha$ -VIII were easily identified as motif I and motif II respectively. The assignment of other conserved regions to equivalent regions in other methyltransferase classes (see table 5.3) cannot be made from the alignment alone.

It is worth mentioning that the N-terminal halves of the methyltransferases *M.FokI*, *M.StsI*, and *M.LlaI* contain the  $\alpha$ -2 subgroup motifs, while their C-terminal halves belongs to the  $\alpha$ -1 subgroup. This points to the possibility that these double-sized methyltransferases did not arise by gene duplication, but by the fusion of two different methyltransferases, and is consistent with the fact that *M.FokI* contains separate strand-specificity domains for the methylation of the upper and lower strand of its asymmetric target site (Looney *et al.*, 1989).

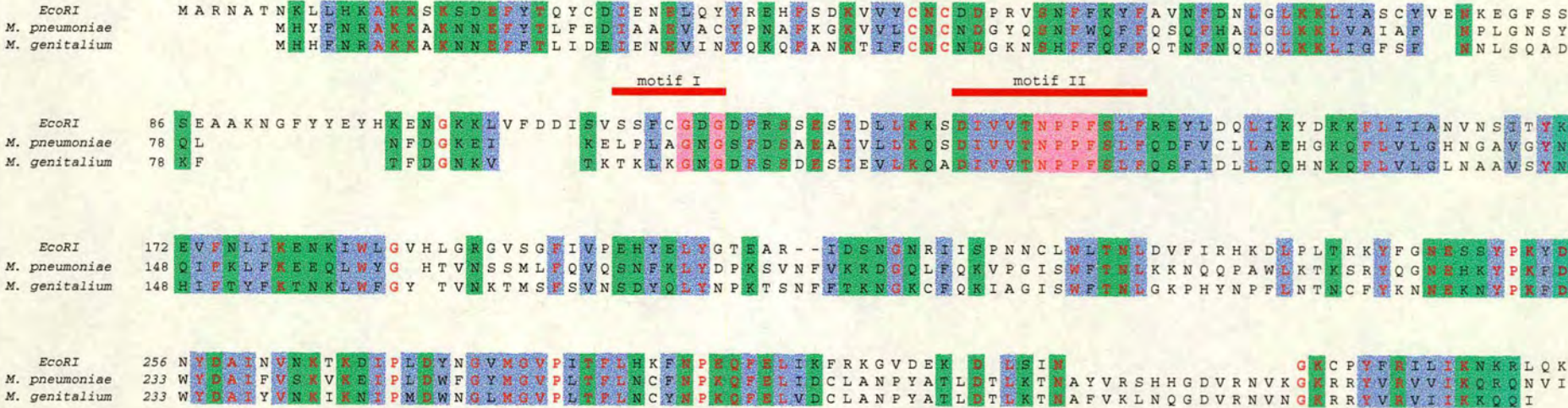
#### 5.2.4.4. *EcoRI* and homologues

Since *EcoRI* methyltransferase does not contain the newly identified conserved regions of the  $\gamma$ -class, I proposed it as the first member of a new class, especially because the distance between motif I and motif II was thought to be too small to accommodate the two conserved regions seen between the two motifs in the  $\gamma$ -class (see Fig. 5.4). This assignment was contested even after the elucidation of the *TaqI* methyltransferase structure had confirmed the importance of the identified conserved regions in the  $\gamma$ -class (Labahn *et al.*, 1994), and alternative alignments produced which put motif I further towards the N-terminus in order to give sufficient space for the placement of the other motifs (Malone *et al.*, 1995).

The identification of highly similar homologues to the *EcoRI* methyltransferase in *Mycoplasma genitalium* (Fraser *et al.*, 1995) and *Mycoplasma pneumoniae* (Himmelreich *et al.*, 1996) finally confirmed that these three methyltransferases belong to a separate class. However the high degree of similarity between them makes it impossible to decide which conserved residues or even which regions are involved in either the methylation or the binding of DNA.



Fig. 5.4: Alignment of *M.EcoRI* and homologues



Legend: Alignment:   Highly conserved motif   Hydrophobic position   Hydrophilic position  
red identical residues



5.2.4.5. RNA methyltransferases

A great number of sequences obtained from the EMBL and other databases turned out to be AdoMet dependent RNA methyltransferases that methylate 16S rRNA, thereby conferring macrolide-lincosamide-streptogramin B resistance to their host. Their abundance can be explained by the fact that most of them are either transposon or plasmid based and thus can easily be passed on to other cells.

The similarity between the RNA methyltransferase sequences ranges from about 20% to nearly 100% identity, making it very easy to align the sequences and identify conserved regions (see Fig. 5.5). The region of similarity shared by all RNA methyltransferases is located near the N-terminus as in the A6 $\gamma$ -class, making it likely that this is the region involved in methylation while the C-terminal region is involved in the binding of the substrate RNA. The similarity of the conserved regions with the conserved regions identified in the A6 $\gamma$ -class - and therefore to the C5- and  $\beta$ -class as well - is striking (see table 5.4), and gives further credibility to the idea of a shared methyltransferase domain.

**Table 5.4: Comparison between some of the motifs identified in RNA methyltransferases with C5- and A6 $\gamma$ -methyltransferases**

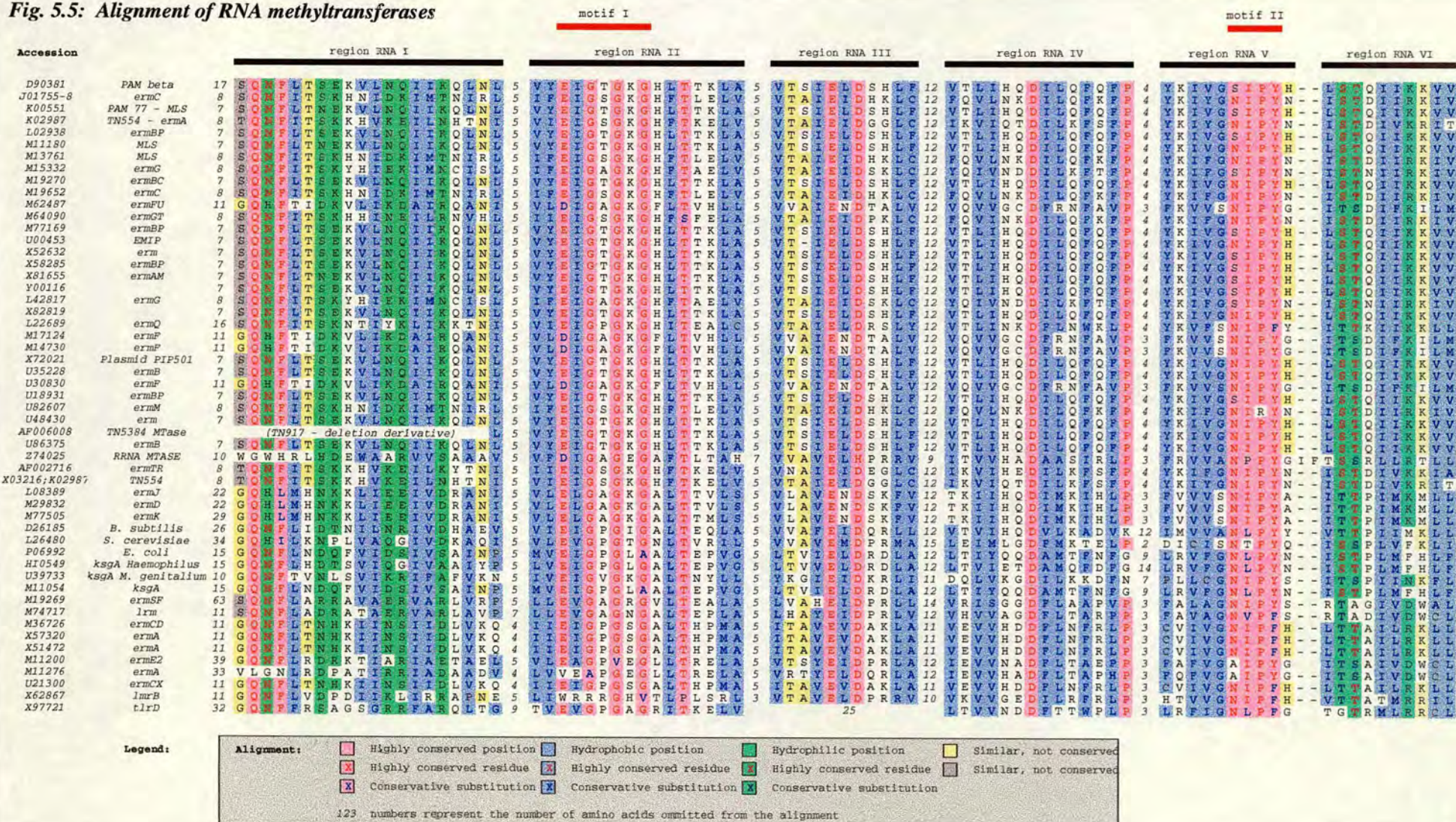
Similar motifs occur in the same order in RNA-, C5- and A6 $\gamma$ -class methyltransferases.

A6 $\gamma$		C5	RNA methyltransferases	
conserved residue	motif	motif	motif	conserved residue
<b>G</b> x <b>F</b> / <b>Y</b> <b>F</b> / <b>Y</b> <b>T</b> <b>P</b>	$\gamma$ -I	C5-X ?	RNA-I	<b>Q</b> <b>N</b> <b>F</b> <b>x</b> <b>t</b> <b>x</b>
<b>x</b> <b>x</b> <b>E</b> / <b>D</b> <b>P</b> <b>x</b> <b>x</b> <b>G</b> <b>x</b> <b>G</b>	$\gamma$ -II	C5-I	RNA-II	<b>x</b> <b>x</b> <b>E</b> / <b>d</b> <b>x</b> <b>G</b> <b>x</b> <b>G</b> <b>x</b> <b>x</b> <b>T</b>
<b>G</b> <b>x</b> <b>E</b> / <b>D</b> <b>x</b> <b>d</b>	$\gamma$ -III	C5-II	RNA-III	<b>a</b> <b>x</b> <b>E</b> <b>i</b> <b>D</b>
<b>x</b> <b>x</b> <b>x</b> <b>x</b> <b>D</b> <b>f</b> <b>x</b>	$\gamma$ -IV	C5-III	RNA-IV	<b>x</b> <b>x</b> <b>x</b> <b>x</b> <b>D</b> <b>x</b> <b>x</b>
<b>D</b> <b>x</b> <b>x</b> <b>x</b> <b>x</b> <b>N</b> <b>P</b> <b>P</b> <b>Y</b>	$\gamma$ -V	C5-IV	RNA-V	<b>x</b> <b>x</b> <b>x</b> <b>x</b> <b>x</b> <b>N</b> <b>I</b> <b>P</b> <b>Y</b>

red: highly conserved residue - blue: hydrophobe residue  
lower case: preference for this residue - bold: similar between A6 $\gamma$  and RNA methyltransferases



**Fig. 5.5: Alignment of RNA methyltransferases**





## region RNA VIII

[illegible]



### 5.2.5. Comparison with structure based sequence alignments

The elucidation and subsequent comparison of crystal structures for *M.HhaI*, *M.TaqI*, and catechol-O-methyltransferase led to the proposal of a universal catalytic domain structure for AdoMet dependent methyltransferases (Schluckebier *et al.*, 1995b). Guided by this common structure a computerised sequence alignment of all the available DNA amino-methyltransferases was performed and conserved motifs specific for each sub-group  $\alpha$ ,  $\beta$ , and  $\gamma$  determined (Malone *et al.*, 1995). Although these alignments are similar to the ones presented in this thesis, some important differences emerged.

- (1) They did not include either Type I, Type III, or RNA methyltransferases, although the similarity between the two Type II and the catechol-O-methyltransferase clearly showed that this domain structure is not confined to Type II enzymes.
- (2) *M.EcoRI* was included in the  $\gamma$ -class and not recognised as separate.
- (3) *M.EcaI* (a  $\beta$ -class enzyme) was wrongly identified and aligned as an  $\alpha$ -class enzyme (after discussion with Xiaodong Cheng their assignment was changed before publication).
- (4) They separated the  $\beta$ -class into A6- and C4-methyltransferases and aligned them differently (at motif VII), failed to spot the first conserved region (motif III), and determined different regions as conserved (motif III and VIII). The elucidation of the *M.PvuII* crystal structure confirmed the alignment presented in this thesis.
- (5) They did not separate the  $\alpha$ -class into two subgroups, and consequently produced a different alignment. While they identified four of the five regions common to all  $\alpha$ -class enzymes (regions  $\alpha$ -I, -II, and -III near the N-terminus and the ubiquitous motif DPPY in region  $\alpha$ -VIII), they failed to spot region  $\alpha$ -XII. Their alignment differs greatly in the remainder of the sequence where it shows only weak conservation.



### 5.2.6. Conclusion

The alignments presented here were the basis for subsequent secondary structure predictions and the structural modelling of part of the HsdM subunit of the *EcoKI* methyltransferase. The elucidation of the *M.TaqI* and *M.PvuII* crystal structures confirms their high accuracy for both the  $\gamma$ - and  $\beta$ -class.

It was quite unexpected to find the same conserved residues not only in Type I, Type II, and Type III enzymes methylating adenine, but among C5-, A6 $\gamma$ -, A6 $\beta$ -, C4 $\beta$ -, and even RNA-methyltransferases, indicating a domain common to the different types of methyltransferases.

## 5.3. Secondary Structure Predictions

### 5.3.1. Chou-Fasman

Over 20 different methods have been proposed for secondary structure predictions (Schulz, 1988; Branden and Tooze, 1991), with the Chou-Fasman method probably being the most widely used. This method is based on a statistical analysis of an x-ray structure database which showed that some amino acids are more commonly found in a particular secondary structure than others. It is therefore possible to assign an average probability to each residue for it being in an  $\alpha$ -helix,  $\beta$ -strand, or in a loop.

When positional effects (like differences in the frequency of residues at a given position in loop regions) and the effects of neighbours in  $\alpha$ -helical and  $\beta$ -sheet structures are taken into account, then the overall accuracy is generally around 50%. However since a large fraction of the errors occur at the end of  $\alpha$  helices and  $\beta$  sheets whereas their central regions are correctly predicted, or arise because of difficulties in distinguishing at times between  $\alpha$  helices and  $\beta$  sheets when both have similar probability values, these predictions still yield very useful results.



### 5.3.2. PredictProtein

It was a natural step forward to use the information from multiple related sequences to increase the accuracy of secondary structure predictions (Rost and Sander, 1993). By aligning the sequences and calculating the probability for a given *position* over all residues, the accuracy could be increased to over 70% overall. It cannot be emphasised enough that the accuracy of the secondary structure prediction is critically dependent on the quality of the alignment.

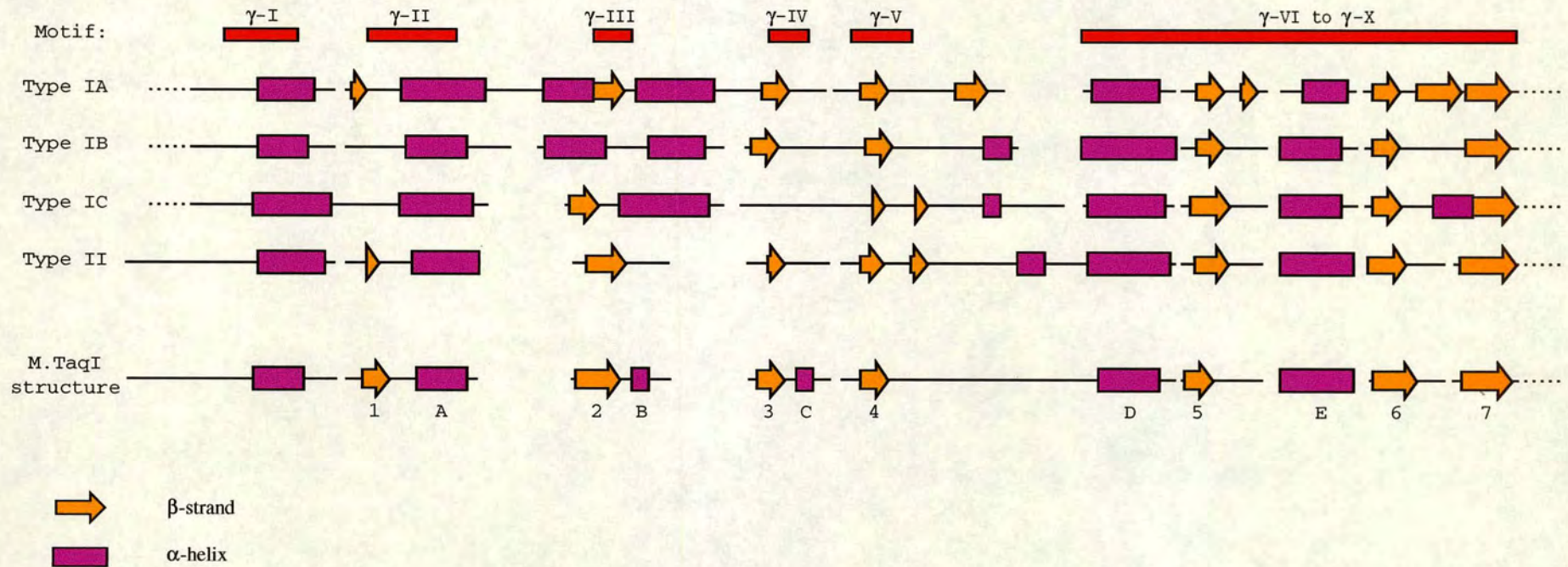
After this new method was made available via the WorldWideWeb, it quickly replaced Chou-Fasman as the method of choice for secondary structure predictions in this work. When only the methyltransferase sequences were submitted for prediction using the PHD mail server (Rost *et al.*, 1994), the alignment method used by the PredictProtein program failed as expected (judged by its failure to align the sequences at motifs I and II, respectively), and the output produced was discarded.

When however the alignment presented in this thesis was submitted for prediction, the output produced showed a remarkable similarity with the actual *M.TaqI* secondary structure as seen in the crystal (see Fig. 5.6), further strengthening the case for a methyltransferase domain common to both Type I and Type II enzymes. It is important to note that the *M.TaqI* structure was not in the database at this time, since the PredictProtein program performs an automated search for and comparison with known protein structures.



### Fig. 5.6 Secondary Structure Prediction

Secondary structure predictions for the region of similarity between Type I families and the Type II  $\gamma$ -class, and comparison with the secondary structure observed in the catalytic domain of the M.*TaqI* crystal structure.



The strands and helices in the M.TaqI structure are numbered as in Schluckebier *et al.*, 1995.



## 5.4. Model Building

The publication of the *M.HhaI* crystal structure had opened the possibility of modelling part of the HsdM subunit of *EcoKI*, but the two sequences were too dissimilar to be aligned over the whole conserved region. But while HsdM could not be aligned with *M.HhaI*, it could be aligned with *M.TaqI*, which in turn could be aligned with *M.HhaI* by employing the fact that motif I and motif II are nearly evenly spaced apart in the two sequences. Before the modelling was attempted however the *M.TaqI* crystal structure was solved and made available to us (Prof. Saenger & Gerd Schluckebier, personal communication).

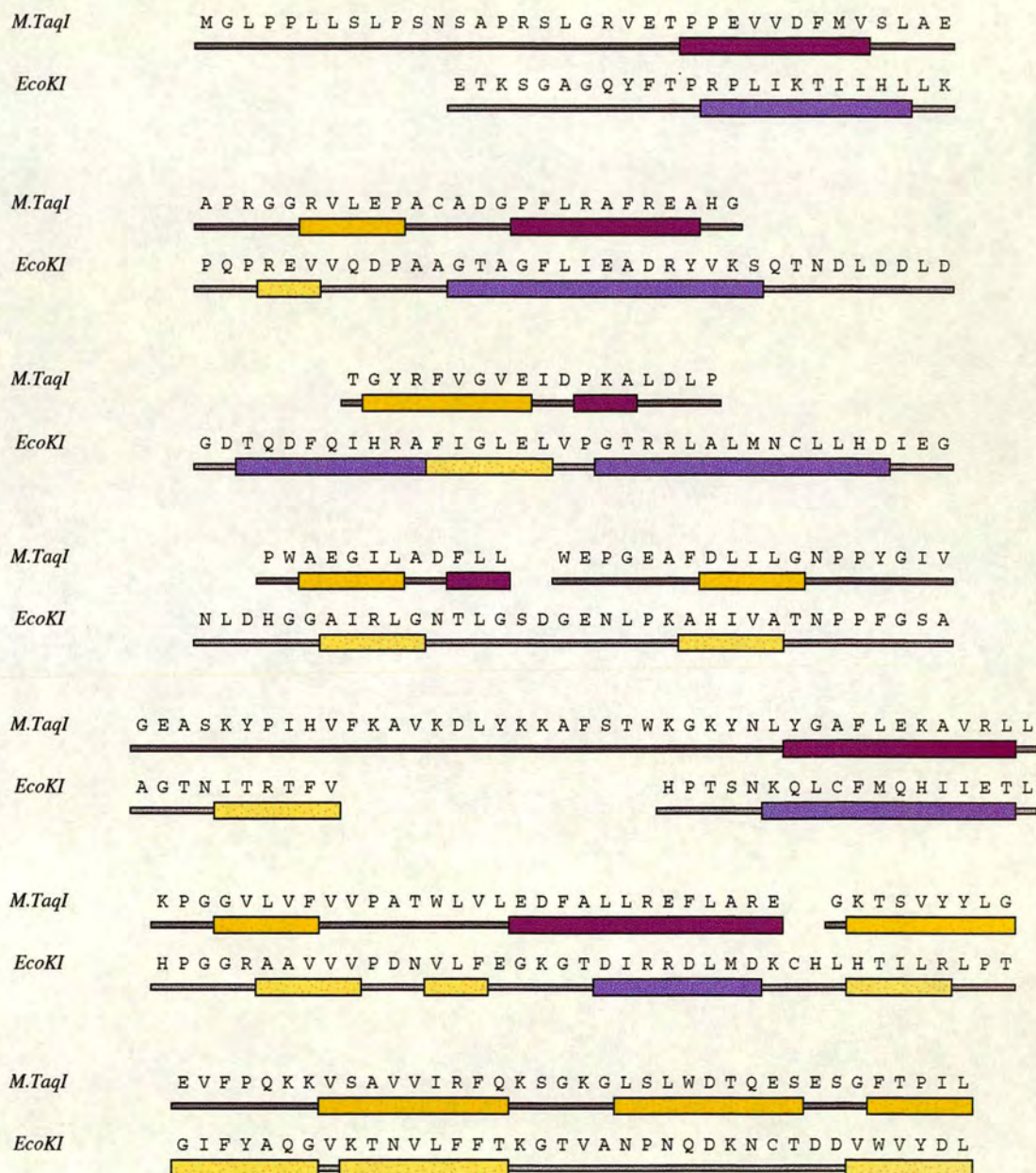
*M.TaqI* and the HsdM subunit of *EcoKI* were aligned over the whole region of similarity (see Fig. 5.7), and this alignment used as basis for the stepwise alteration of the *M.TaqI* sequence into the *M.EcoKI* sequence. By successively exchanging one amino acid residue in the *M.TaqI* structure against the corresponding amino acid residue in *M.EcoKI*, the crystal structure was effectively mutated into the other sequence.







Insertions and deletions in *M.EcoKI* occur on surface loops in the *M.TaqI* structure, as would be expected if the two enzymes share a common domain. While excess loops in the *M.TaqI* structure (equal to a deletion in *M.EcoKI*) could easily be deleted, insertions into the *M.TaqI* structure needed to be modelled. Shane Sturrock, who did all the molecular modelling using the FRODO and SYBIL (by Tripos) molecular modelling programs, wrote a special program to search the structural databases for similar loops that could be used as starting points for the modelling of the insertions.

When the *M.TaqI* structure had finally been converted into the *M.EcoKI* sequence and the resulting structure undergone several rounds of energy minimisation, the end result (see Fig. 5.8) was a stable structure without any steric conflicts (Shane Sturrock, personal communication).



**Fig. 5.7: Alignment *M.TaqI* and *M.EcoKI***



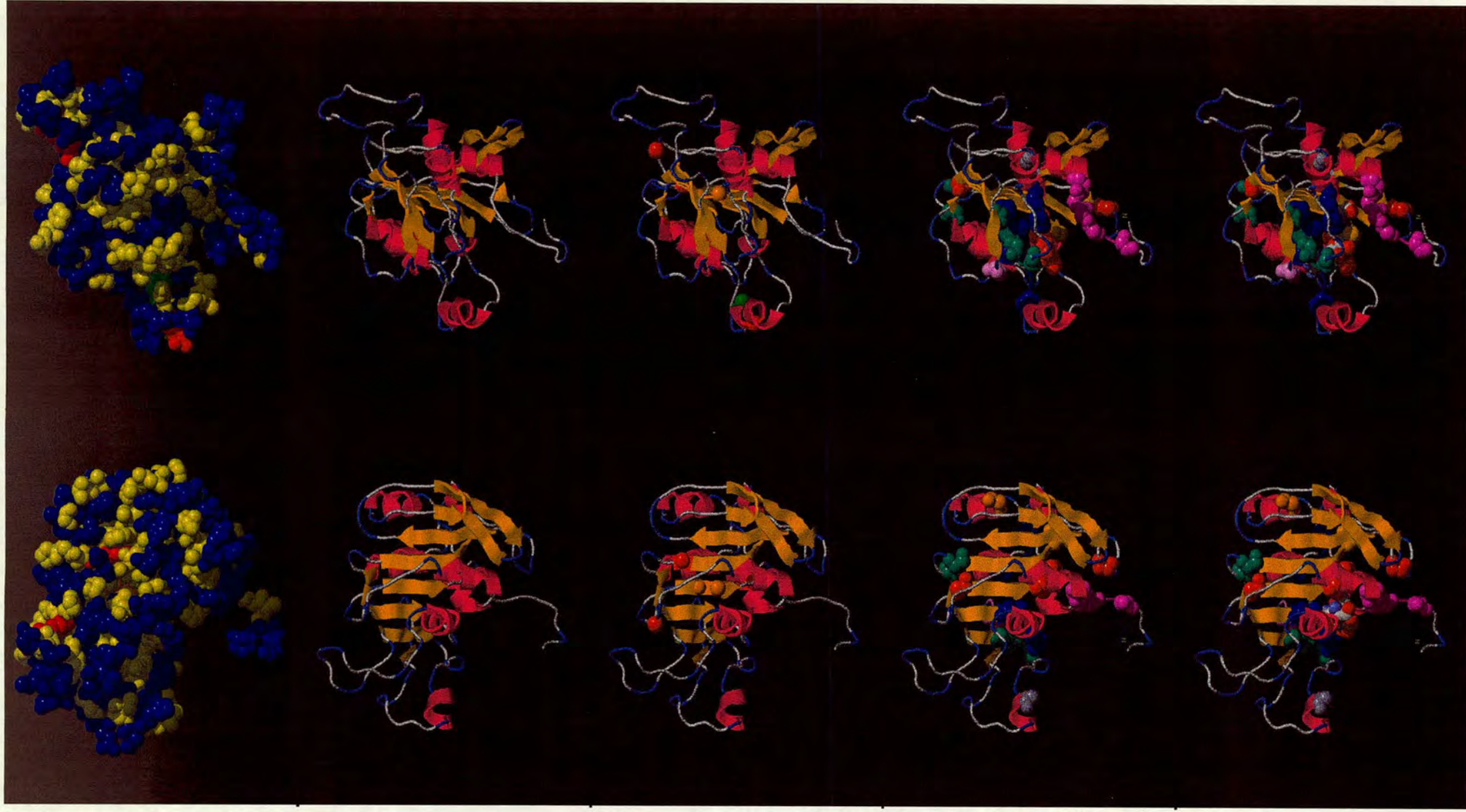
M. <i>TaqI</i> crystal structure		<i>EcoKI</i> prediction	
	α-Helix		α-Helix
	β-Sheet		β-Sheet
	Loop		Loop



**Fig. 5.8: Model of the M.EcoKI methyltransferase domain**

- |   | <i>side view</i>  | <i>front view</i>      |
|---|---|------------------------|
| a) <b>space filling model of the M.EcoKI methyltransferase domain</b>   |   |                        |
| blue:   | polar amino acids   |                        |
| yellow:   | hydrophobic amino acids   |                        |
| red:  | Trypsin cleavage sites (Arg168, Arg279, Arg305)                   |                        |
| green:  | Chymotrypsin cleavage site (Phe281)                               |                        |
| b) <b>cartoon representation of the M.EcoKI methyltransferase domain</b>  |   |                        |
| magenta:  | $\alpha$ -helices   |                        |
| yellow:   | $\beta$ -sheets   |                        |
| pale blue:  | turns   |                        |
| white:  | others  |                        |
| c) <b>cartoon representation of the M.EcoKI methyltransferase domain showing the protease cleavage sites</b>  |   |                        |
| magenta:  | $\alpha$ -helices   |                        |
| yellow:   | $\beta$ -sheets   |                        |
| pale blue:  | turns   |                        |
| grey:   | others  |                        |
| red:  | Trypsin cleavage sites (Arg168, Arg279, Arg305), C $\alpha$ -atom |                        |
| green:  | Chymotrypsin cleavage site (Phe281), C $\alpha$ -atom             |                        |
| orange:   | Elastase cleavage site (Val263, Ala264), C $\alpha$ -atom         |                        |
| d) <b>cartoon representation of the M.EcoKI methyltransferase domain, showing the conserved regions (backbone atoms of highly conserved residues are shown in space filling representation)</b> |   |                        |
| magenta:  | $\gamma$ -I   | mid-blue: $\gamma$ -VI |
| blue:   | $\gamma$ -II  | green: $\gamma$ -VII   |
| green-blue:   | $\gamma$ -III   | yellow: $\gamma$ -VIII |
| violet:   | $\gamma$ -IV  | orange: $\gamma$ -IX   |
| red:  | $\gamma$ -V   | brown: $\gamma$ -IX    |
| grey:   | the asparagine highly conserved among Type I methyltransferases   |                        |
| e) <b>as d), with AdoMet modelled into the structure</b>  |   |                        |





a)

b)

c)

d)

e)



## 5.5. How do other data fit into the model?

The degree of success in modelling a structure is not assessed by how pretty it looks or how similar it is to the template used, but by its internal and external consistency. The internal consistency can be calculated by the molecular modelling software used, and is high for the methyltransferase domain model since it resulted in a stable structure with no steric conflicts (Shane Sturrock, personal communication). The external consistency can be assessed by how well the model agrees with experimental data.

Previous investigations of the *EcoKI* methyltransferase structure were done by limited proteolysis experiments, and the cleavage sites for some of the fragments determined by N-terminal sequencing (Cooper and Dryden, 1994). It is important to bear in mind that the identification of cleavage sites depends on the accumulation of intermediates in the proteolytic pathway. It does not require the cleavage site to be located on the surface of the protein, since some of the proteolysis sites may have become accessible only after other sites had been cleaved.

It is also possible that the two HsdM subunits in *M.EcoKI* follow different proteolytic pathways despite being identical. Since the interface to the HsdS subunit is likely to have different surroundings depending on which target recognition domain is located close by, access to proteolytic sites might be blocked for one HsdM subunit, but not the other.

Several cleavage sites are located on exposed loops in the model (*e.g.* Arg168, Arg279, Phe281; see Fig. 5.8) and therefore easily accessible to the protease. The presence of the cofactor AdoMet prevents cleavage at the exposed loop containing Arg279 and Phe281, consistent with the close proximity of the cofactor-binding site to the loop and suggesting a conformational change similar to the one observed for an equivalent loop in *M.HhaI* (Klimasauskas *et al.*, 1994).

In the absence of AdoMet the whole HsdM subunit is far less stable, and especially the N-terminal half containing the AdoMet binding site is rapidly degraded, providing access to cleavage sites usually buried in the interior of the protein (*e.g.* Val263, Ala264, Arg305). The C-terminal half is more resilient, as can be seen by the accumulation of C-terminal fragments which are only slowly digested by trypsin and chymotrypsin, and even the faster digestion by elastase shows the transient accumulation of a C-terminal fragment.



A mutational analysis of the *M.EcoKI* methyltransferase domain has so far been restricted to the ubiquitous motifs I and II (Willcock *et al.*, 1994). However homologues of a protein are naturally occurring and usually fully functional mutants, and every sequence alignment can therefore be regarded as a collection of mutations and be used for mutational analysis.

Mutations in the core of a protein are likely to destabilise its structure and would therefore be expected to be rather rare and conservative in nature, while mutations at the periphery are more easily accommodated and should consequently be more common and varied. Positions with a high degree of sequence variability are therefore likely to be located in loop regions or on the surface of a protein (although one should bear in mind that this does not apply to receptor proteins and protein-protein interfaces).

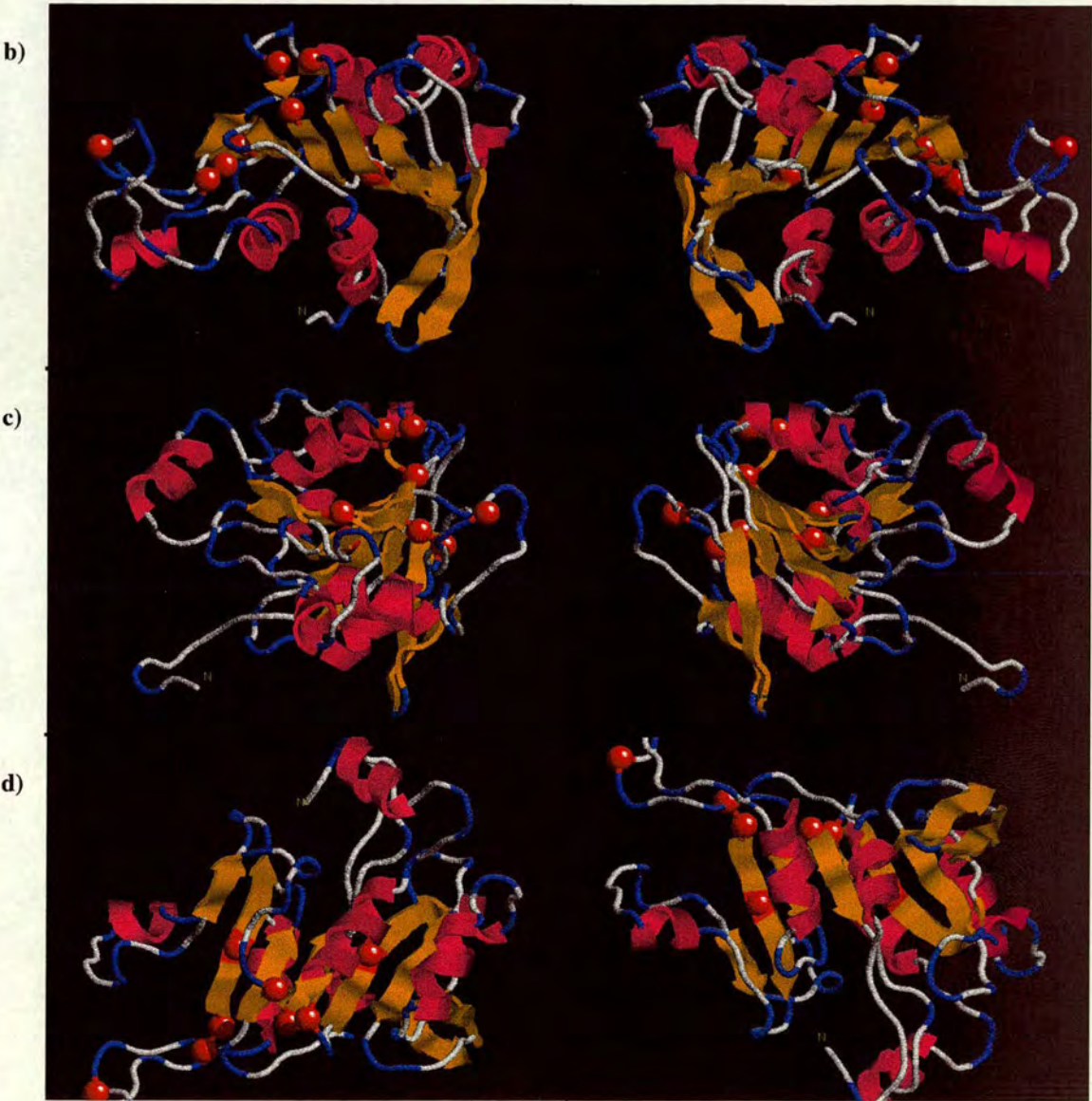
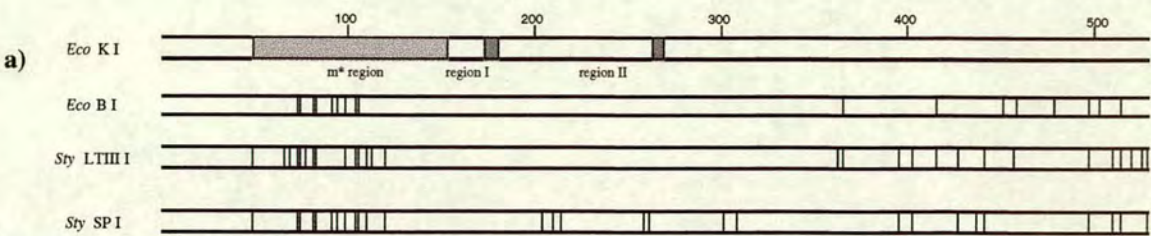
Previous sequence comparisons among Type IA systems left out the analysis of HsdM sequences, since they were not expected to yield any useful information due to the near identity of the HsdM homologues with each other. However if one ignores the number of differences between the sequences and looks at their distribution instead, a comparison between the *M.EcoKI*, *M.EcoBI*, and *M.StyLTIII* hsdM sequences shows that all the differences lie in the N- and C-terminal thirds, while the central parts are identical. This apparently non-random distribution (see Fig. 5.9) was the first indication that the central part - which includes motif I and motif II - might be a domain, and was also consistent with the location of the m\* region in the N-terminal third.

The less similar *StySPI* sequence contains differences in the central part, and can therefore be used to test the prediction on the modelled structure. As expected the majority of differences between the sequences are found in loop regions (6 differences among 144 residues), while only highly conservative substitutions take place in the core (2 in 96 residues). It is therefore twice as likely for a difference to be located in a loop region than within the conserved structure.

Even more significant is the strong asymmetry seen in the spatial distribution of the differences in the model (see Fig. 5.9). This asymmetry strongly indicates that the sides of the domain without differences might be involved in domain, subunit, or protein-DNA interactions, while the side containing all the differences would point outwards away from the complex.



**Fig. 5.9: Location of differences between Type IA sequences**



a) Location of sequence differences between Type IA sequences

b) side view  
c) back  
d) top

other side  
front  
bottom

the location of different amino acids are depicted red in space filling representation



The modelled domain also shows plausible locations for mutations in motif I and motif II that affect the AdoMet binding or catalytic activities of *M.EcoKI* (Willcock *et al.*, 1994), and is consistent with the high conservation of these motifs in all DNA methyltransferases.

The construction of pT7-7-MS( $\Delta m^*$ ) not only deleted the  $m^*$  region from HsdM, but led to the replacement of amino acids which form part of the highly conserved motif  $\gamma$ -I (see Chapter 3, table 3.1). The accumulation of HsdS (but not of the truncated HsdM) seen after induction of this construct can be accounted for by assuming a destabilising effect of these substitutions on the HsdM subunit, leading to its proteolysis in the cell. The remaining HsdS could accumulate and bind to cellular DNA, thereby inhibiting the cells growth as observed during induction.

## 5.6. Proposing a new structure for *M.EcoKI*

The existence of a distinct methyltransferase domain in *M.EcoKI* has important implications for the structure of the HsdM subunit. Results from limited proteolysis experiments had suggested a two-domain model in which an N-terminal domain containing the  $m^*$  region and motif I is connected via a short loop containing motif II to a C-terminal domain that contacts the HsdS subunit (see Willcock *et al.*, 1994).

The data presented here are consistent with the location of motif II in a flexible loop. However this loop does not connect two domains with each other, but is part of the methyltransferase domain. Consequently the N-terminal  $\approx 140$  amino acids containing the  $m^*$  region are likely to form a separate domain, as are the  $\approx 150$  amino acids at the C-terminus which encompass a region implicated in protein-protein interactions with the HsdS subunit (see Fig. 5.10).

By using the available structural information for both the modelled *M.EcoKI* methyltransferase domain and the *M.HhaI* crystal structure in complex with DNA, and combining it with data indicating a symmetrical structure for Type I DNA methyltransferases (see Kneale, 1994 and model therein), D. Dryden and S. Sturrock constructed a composite model of the *EcoKI* methyltransferase on DNA (see Fig. 5.11, Dryden *et al.*, 1995). This model is in excellent agreement with all the available biochemical data including DNA footprinting (Powell and Murray, 1995), protein-DNA crosslinking (Chen *et al.*, 1995), and small angle X-ray scattering

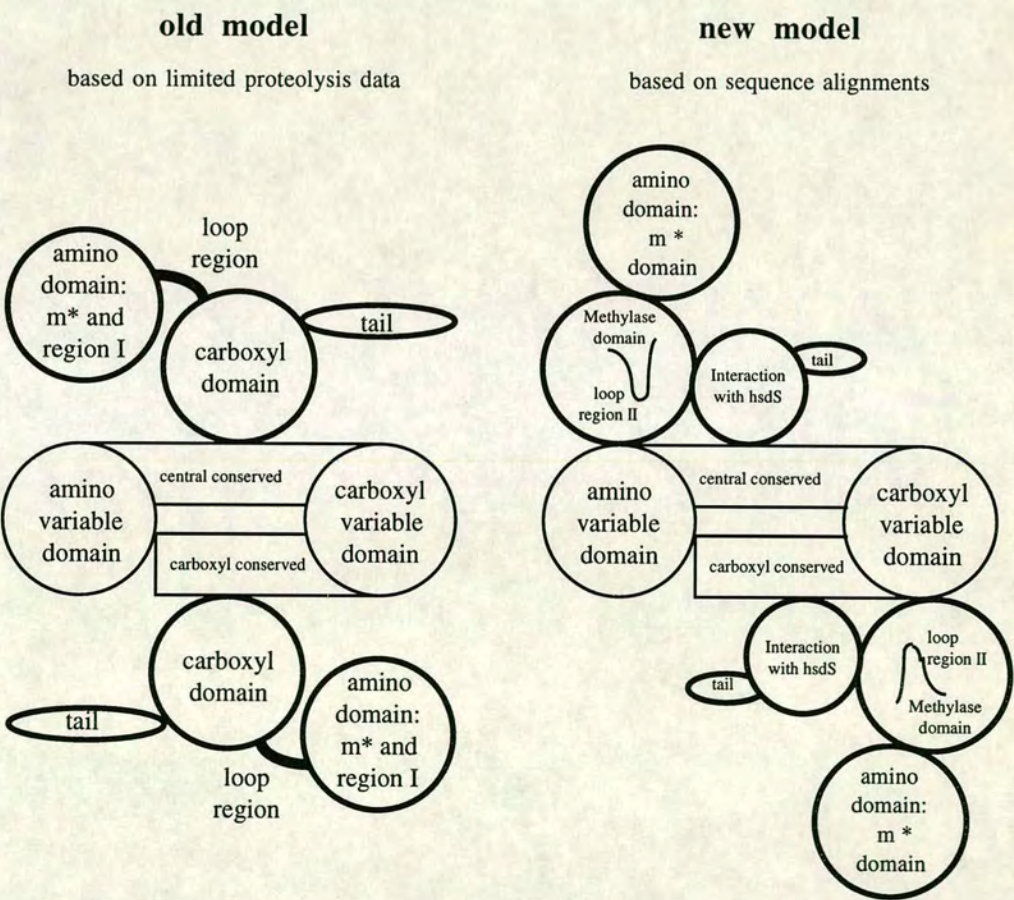


experiments (Taylor *et al.*, 1994), as well as measurements of the hydrodynamic shape of the protein (Powell *et al.*, 1993; Taylor *et al.*, 1994).

When the differences in sequence between different Type IA methyltransferases were located in the modelled composite structure, they were found to lie on the outside of the complex as expected.

**Fig. 5.10: Proposed model for the EcoKI methyltransferase**

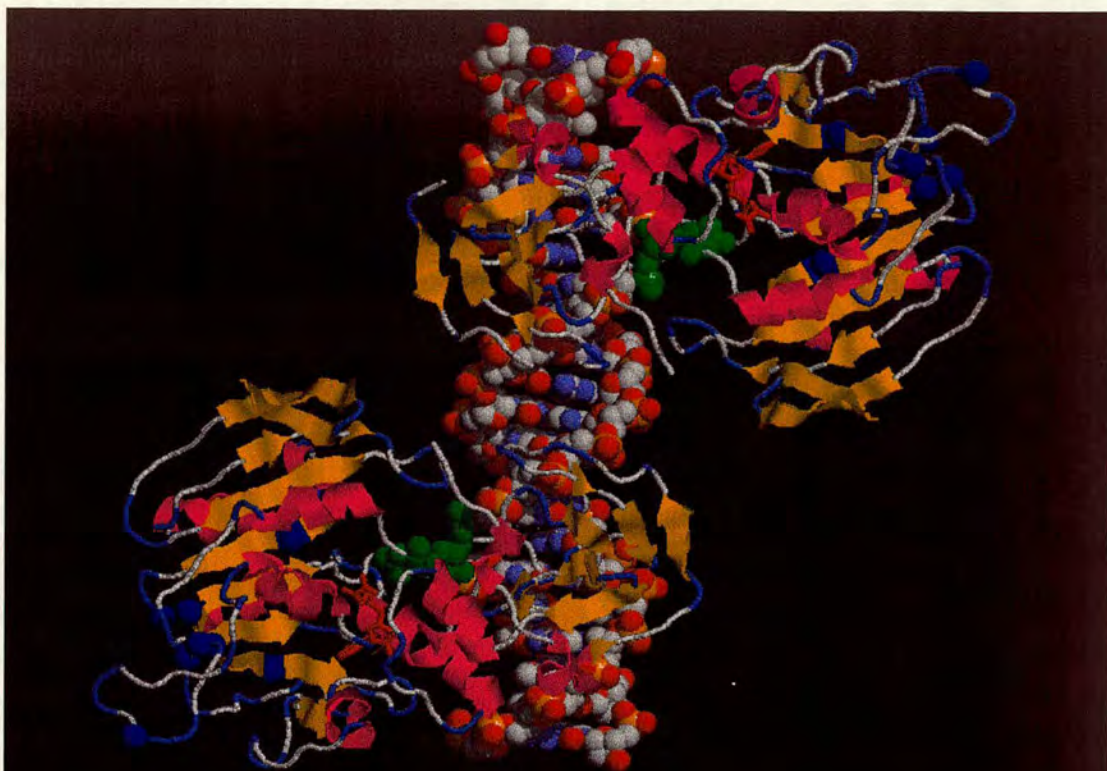
bold lines: HsdM      thin lines: HsdS



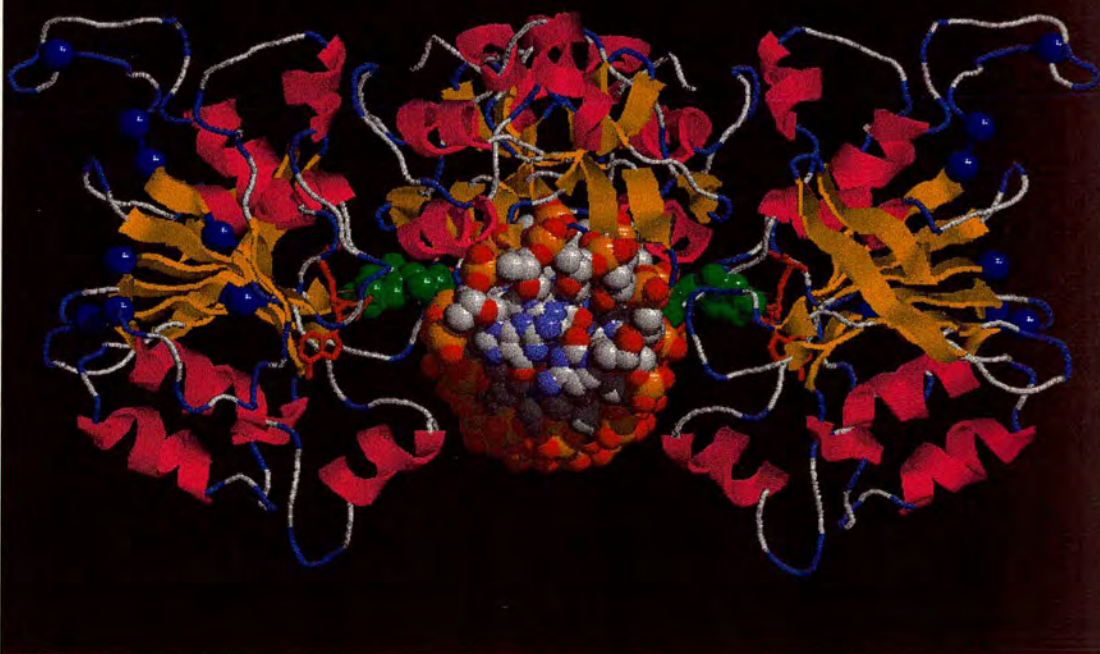


**Fig. 5.11: Model of the *Eco* KI methyltransferase on DNA**

a)



b)



- Protein: cartoon representation  
 magenta:  $\alpha$ -helices  
 yellow:  $\beta$ -sheets  
 pale blue: turns  
 white: others  
 blue: location of differences, C $\alpha$ -atom shown in space filling representation
- DNA: space filling representation
- AdoMet: stick representation, coloured red
- target base: space filling representation, coloured green



**CHAPTER 6:**

**CONCLUSIONS**

**AND GENERAL**

**DISCUSSION**



The structure of Type I methyltransferases has been postulated to be symmetrical by way of a "circular" organisation of the HsdS subunit with the two TRDs spaced appropriately by the conserved regions (see fig. 1n; Meister *et al.*, 1993; Kneale, 1994). Methylation interference (Powell and Murray, 1995) and UV crosslinking experiments (Chen *et al.*, 1995) are consistent with the two TRDs of HsdS contacting the major groove of the DNA at their respective half sites of the bipartite recognition sequence, with the conserved regions of HsdS spanning the intervening minor groove. The HsdM subunits are placed on either side of the conserved regions of the HsdS subunit, one interacting with the repeat sequences in the central conserved region and the other with the amino- and carboxyl conserved region of HsdS.

The two parts of the bipartite recognition sequence of Type I enzymes are separated by either 10 or 11 base pairs, and therefore the N6 positions of the substrate adenines project into the major groove on the same face of the DNA helix, on either side of the minor groove (see fig. 1o; Burckhardt *et al.*, 1981a; Bickle, 1987). The spatial difficulties of both sequence recognition by HsdS and methylation by HsdM in the major groove may be reconciled if the *EcoKI* methyltransferase can flip its target adenines out of the DNA helix in the same way as the *HhaI* C5 cytosine methyltransferase (Klimasauskas *et al.*, 1994). The similarities between Type I and Type II systems revealed in chapter 5 strongly support the existence of a common methylation domain and therefore agree with a common base flipping mechanism as well.

What are the implications of a common methylase domain for the subunit structure of HsdM? HsdM must have at least three domains, with the methylase domain being the central one. The methylase domain binds AdoMet and has to be located in a position close to the target base for methylation. The C-terminal domain is involved in protein:protein interactions with the HsdS subunit (Cooper and Dryden, 1994), presumably with the conserved regions spanning the minor groove of the DNA, and is therefore positioned accordingly. The location of the start of the modelled domain suggests that the N-terminal m\* domain is folded up against the back of the methylase domain, alongside the C-terminal domain.

One of the main differences between Type IA and Type II systems is that methylation in Type IA systems is dependent on the presence of a methylated base at the other half site of the recognition sequence. Contrary to Type II systems they therefore have to probe the methylation state at one half site and relay this information to the other half site before methylation can proceed. Mutations in the N-terminal third



of HsdM that eliminate the dependency of methylation on the presence of a methylated base at the other half target site therefore suggest that the N-terminal domain of HsdM is involved in the relay mechanism.

As with every piece of data one has to question whether what someone sees is the cause, or an effect of the phenomenon under investigation. A decrease in the binding affinity for AdoMet would be expected to lead to a decrease in methylation, not to an increase in the methylation activity on unmethylated DNA as seen for the  $m^*$  mutants. However the mutants L113Q and L134V bind AdoMet significantly worse than the wild-type methyltransferase. But the data from the ANS titration clearly indicate that only one of the two sites present in the methyltransferase is affected, while the other binds AdoMet with the same  $K_d$  as the wild-type. Since the two HsdM subunits despite being identical are located near different HsdS target recognition domains it is conceivable that the mutation in HsdM might lead to a reduction in the  $K_d$  for AdoMet in one of them but not the other (possibly due to steric interferences with HsdS). However if this were the case one would expect methylation at one target site to be significantly worse than at the other, and the results from the methylation assays clearly show that L113Q methylates both hemi-methylated substrates with equal efficiency. Both HsdM subunits therefore are equally methylation competent, both carry the same mutation, and yet one site displays a higher  $K_d$  for AdoMet.

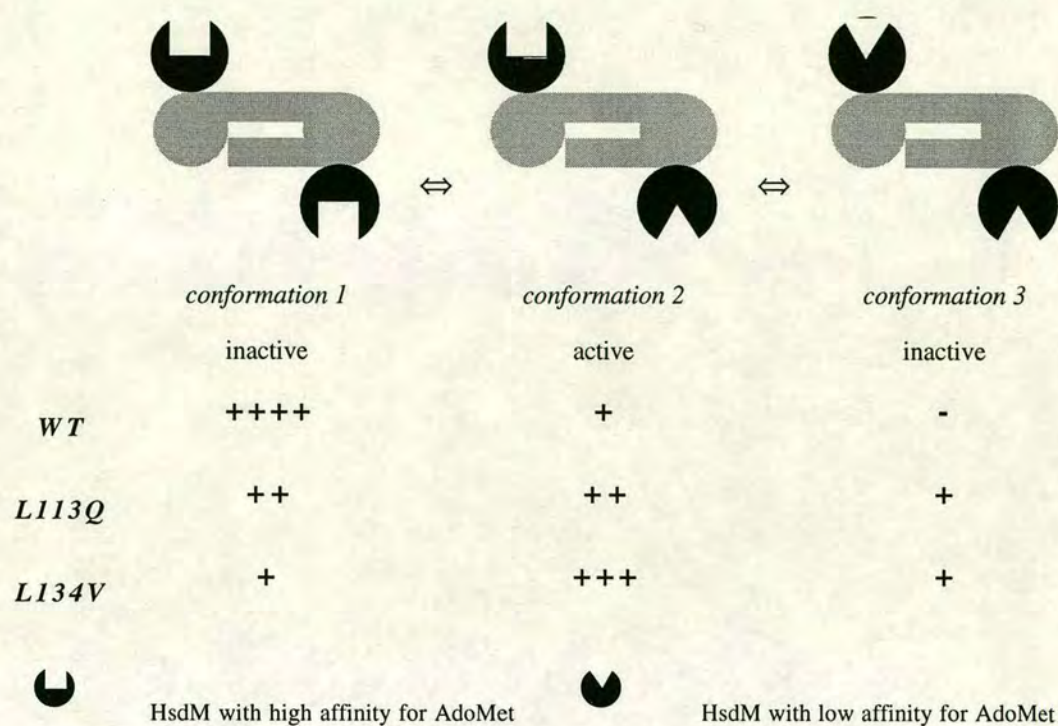
However the data do not say that it is always the same site that binds AdoMet with a higher  $K_d$ . Since both HsdM subunits are identical it is more likely that the methyltransferase adopts a conformation where one site has a high affinity for AdoMet, while the other one has a low affinity. Consequently a lower overall affinity for AdoMet would be displayed by the complete enzyme. The measured  $K_d$  therefore depends on how likely it is for the HsdM subunit to adopt one conformation or the other and therefore what fraction of the HsdM subunits is binding AdoMet with high affinity (see fig. 6.1).

What relevance does the existence of two conformations with different binding affinities for AdoMet in the mutants have for the reactions of the wild-type enzyme? And most important, what triggers the conformational change?

The most likely candidate for the trigger is the target base. It has been suggested that *EcoKI* uses AdoMet to probe the methylation state of its target site, but instead of moving AdoMet towards the DNA as suggested it seems likely that the target base is swung out as seen for *M.HhaI*. If the base is unmethylated then it will be in close proximity to the methyl group of AdoMet, ready for methylation. If however the base



is already methylated, then it will induce the transition of the methylation domain to a conformation with lower affinity for AdoMet, possibly in order to avoid a steric conflict between the methyl group of the base with the methyl group of the bound AdoMet.



**Fig. 6.1: Proposed conformational states of the *EcoKI* methyltransferase**

While the wild-type methyltransferase is mainly present in conformation 1, thereby showing high affinity for AdoMet, the mutant HsdM subunits are more likely to adopt the conformation with low affinity for AdoMet, resulting in a lower apparent affinity and higher apparent  $K_d$ . The numbers of + indicate the estimated preference of wild-type and mutant methyltransferases for each conformation.

*EcoKI* possesses two HsdM subunits, and information about the methylation state of their respective target base has to be relayed to the other subunit. A conformational change would be the most obvious way of doing it. On unmodified DNA AdoMet is bound at both HsdM subunits in the same way. On hemi-methylated DNA however the methylated base induces a conformational change in the HsdM subunit it is bound by. This conformational change opens the way for methylation at the other target site to proceed, and could either be achieved by the removal of a steric block or by direct activation. With fully methylated DNA both HsdM subunits adopt the conformation with lower affinity for AdoMet.



This model immediately suggests how the  $m^*$  mutant methyltransferases methylate unmodified DNA with increased efficiency (see fig. 6.1). If they are more likely to adopt the conformation that triggers methylation at the other target site because of their mutation, then methylation at the other site will mistakenly be instigated even in the absence of a methylated base as trigger. Consequently in the restriction competent complex of the methyltransferase with HsdR the  $m^*$  mutations would result in a shift in the balance from restriction towards methylation, as inferred from *in vivo* studies by Kelleher (*et al.*, 1991).

What is the nature of the conformational change and how is the information relayed to the other HsdM subunit? It seems unlikely that the HsdS subunit is involved in the process since TRDs of different systems and even whole HsdS subunits can be swapped with each other. The sole function of the HsdS subunits therefore seem to be to correctly position the HsdM subunits on the DNA for methylation of the appropriate target sequence. Since the two methylase domains are too far apart to directly interact with each other, the obvious candidate to relay the information to the other HsdM subunit is the  $m^*$  domain. In this model (see fig. 6.2) the flipped-out methylated target base would induce a conformational change in the methylase domain which not only lowers its affinity for AdoMet, but in turn changes its interaction with the  $m^*$  domain. This leads to a repositioning of the  $m^*$  domain that enables methylation at the other target side to proceed. That a proteolytic site near the C-terminus of HsdM is less available in the presence of hemi-methylated DNA (Cooper and Dryden, 1994) together with the positioning of the domains relative to each other in the *EcoKI* model lends support to this model.

The  $M_2S_1$  complex of *EcoKI* is inactive on unmodified DNA. And while  $M_1S_1$  is able to specifically bind DNA (L. Powell, personal communication), it methylates neither hemi-methylated nor unmodified DNA (D.T.F. Dryden, personal communication), in contrast to Type II methyltransferases which too possess just a single methylase domain. The conclusion must therefore be that methylation not only requires the presence of the second HsdM subunit, but is most likely inhibited in the  $M_1S_1$  complex.

That two of the  $m^*$  mutations are actually located in the modelled methylase domain points to the strong integration of the activities of the  $m^*$  and methylation domains and makes it seem possible that it is the  $m^*$  domain that blocks methylation by the methylase domain. On binding hemi-methylated DNA, the methylated base triggers a conformational change in the methylase domain, which in turn repositions



the m\* domain. This is accompanied by a symmetrical change in the position of the m\* domain in the other HsdM subunit, in accordance with the concerted model for allosteric proteins (see for example the archetypal allosteric protein hemoglobin where binding of ligand at one site has no direct effect on the affinities of the other sites, but alters the conformational equilibrium between two alternative quaternary conformations of the protein; Monod *et al.*, 1965; Rubin and Changeux, 1966). The repositioning of the m\* domain in the other HsdM subunit promotes the adoption of the low affinity conformation of the methylase domain normally adopted with a methylated base, thereby instigating methylation at the other target site.

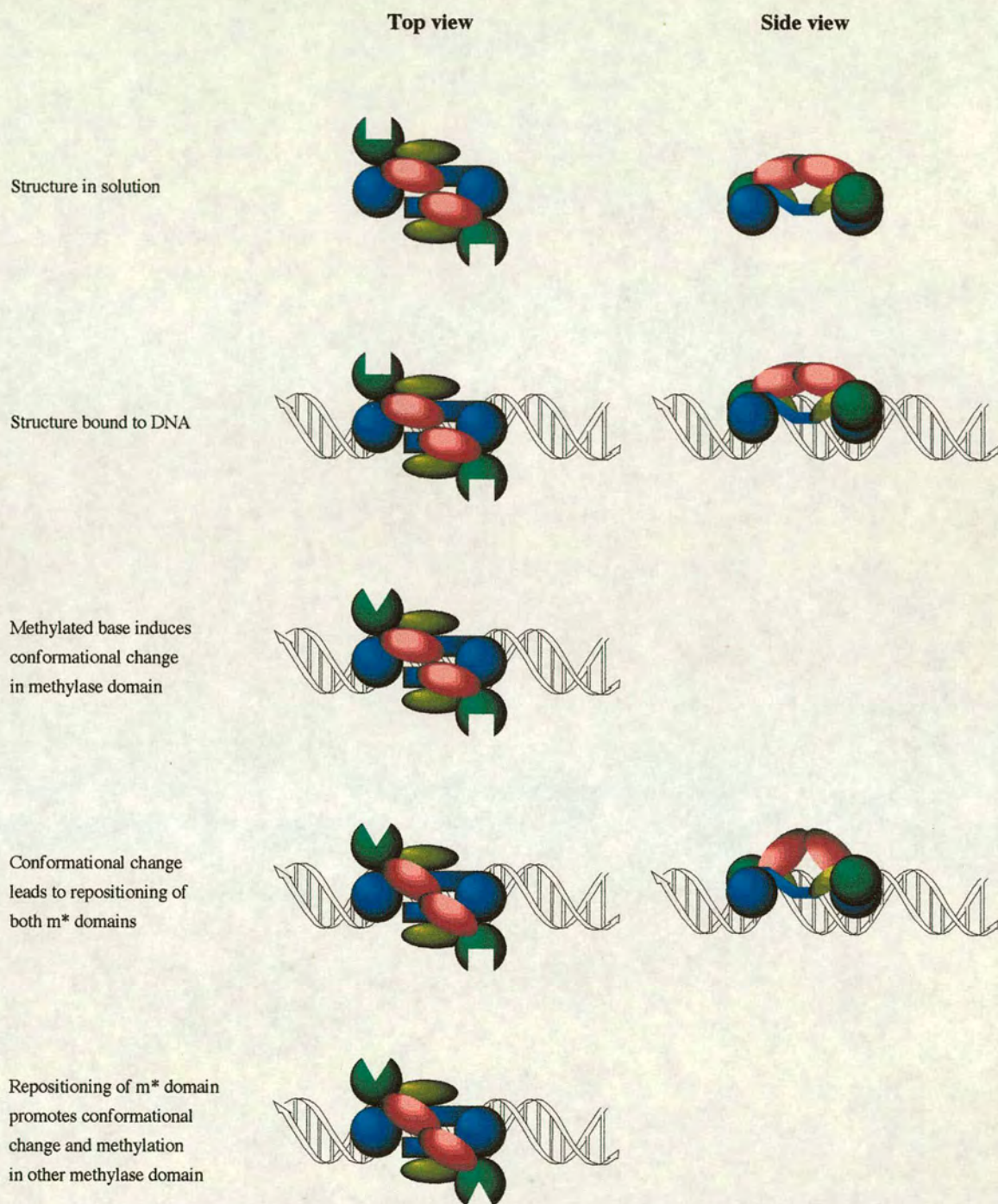
In this model the m\* domains adopt one position on unmodified DNA, and another one on both hemi- and fully methylated DNA. This is borne out by footprinting studies of the M<sub>2</sub>S<sub>1</sub> complex that show the same symmetrical pattern of methylation interference on both hemi- and fully methylated DNA (regardless of the presence or absence of AdoMet), but a different one on unmodified DNA (see fig. 10).

The low level of activity seen in preparations of the M<sub>1</sub>S<sub>1</sub> complex of the wild-type *Eco*KI methyltransferase might be due to contamination by M<sub>2</sub>S<sub>1</sub>, or rearrangements of the subunits in solution to form an active complex of the form M<sub>2</sub>S<sub>1</sub> (D.T.F. Dryden, personal communication). It could however also be the effect of sporadic changes in the conformation of the HsdM subunit, similar to the low level of methylase activity observed for the M<sub>2</sub>S<sub>1</sub> complex on unmethylated DNA. If this is the case, then preparations of M<sub>1</sub>S<sub>1</sub> complexes harbouring one of the m\* mutations might show a marked increase in the level of methylase activity relative to the wild-type enzyme.

It should also be noted that the presence of AdoMet increases the affinity of the *Eco*KI methyltransferase for DNA. Consequently the adoption of the low affinity conformation on fully methylated DNA should result in the release of the cofactor, thereby promoting the dissociation of the DNA-protein complex.

The effects of the anti-restriction system *ral* that promotes the wild-type methyltransferase to methylate unmodified DNA (Zabeau *et al.*, 1980; Loenen and Murray, 1986) could simply be achieved that by binding to the methyltransferase it stabilises or promotes the adoption of the low affinity conformation in HsdM. The failure to see m\* activity in strains overexpressing the *ral* gene (G. King, personal communication) can simply be explained by the fact that an abundance of the Ral





**Fig. 6.2: Proposed mechanism for the methylation reaction of the *EcoKI* methyltransferase**

HsdS: blue - methylase domain; green - HsdM/HsdS interaction domain; yellow-green - m\* domain; red



protein would cause all of the HsdM subunits to adopt the low affinity conformation, effectively mimicking fully methylated DNA.

How can the function of the m\* domain be further investigated? If the m\* domain indeed both blocks methylation by the methylase domain in the same HsdM subunit yet is required to relieve the methylation block in the other HsdM subunit, then the assumptions on the effect of N-terminal deletions in HsdM as outlined in the introduction to chapter 3 do not hold. It also suggests that there may be two different kinds of m\* mutations: one which stimulates methylation by promoting the adoption of the low affinity conformation in the other HsdM subunit, and another one in which the block to methylation has been removed, resulting in methylation by its own HsdM subunit in the absence of activation (the m\* mutations located within the methylase domain, *i.e.* mutations of Ser144 and Arg153, may be likely candidates for this type of action). Consequently these two types can be distinguished by assaying the methylation activity of the M<sub>1</sub>S<sub>1</sub> complex: while the first type requires activation by the other HsdM subunit and should therefore be inactive, the second type should result in an active complex.

Deletion of the m\* domain would have very different effects in these two types of mutants. Deletion of the m\* domain in mutant HsdM of the first type (as well as in wild-type HsdM) would result in an inactive methyltransferase, while deletion of the m\* domain in mutant HsdM of the second type (effectively leaving within the truncated HsdM subunit the mutation removing the block to methylation) would result in an active methyltransferase.

Although the *Eco*KI methyltransferase can be purified in vast quantities, all attempts at crystallisation for subsequent structure determination have failed so far (D.T.F. Dryden, personal communication). Given the apparent multitude of conformational states Type I systems can adopt as evident from several studies (see among others Hadi *et al.*, 1975; Yuan *et al.*, 1975; Burckhardt *et al.*, 1981a; Burckhardt *et al.*, 1981b; Kneale, 1994; Taylor *et al.*, 1994; Mernagh and Kneale, 1996), this result seems hardly surprising. However the information about the domain structure of HsdM presented in this thesis opens the possibility to express and purify the domains separately, making them more malleable to crystallisation and structure determination.

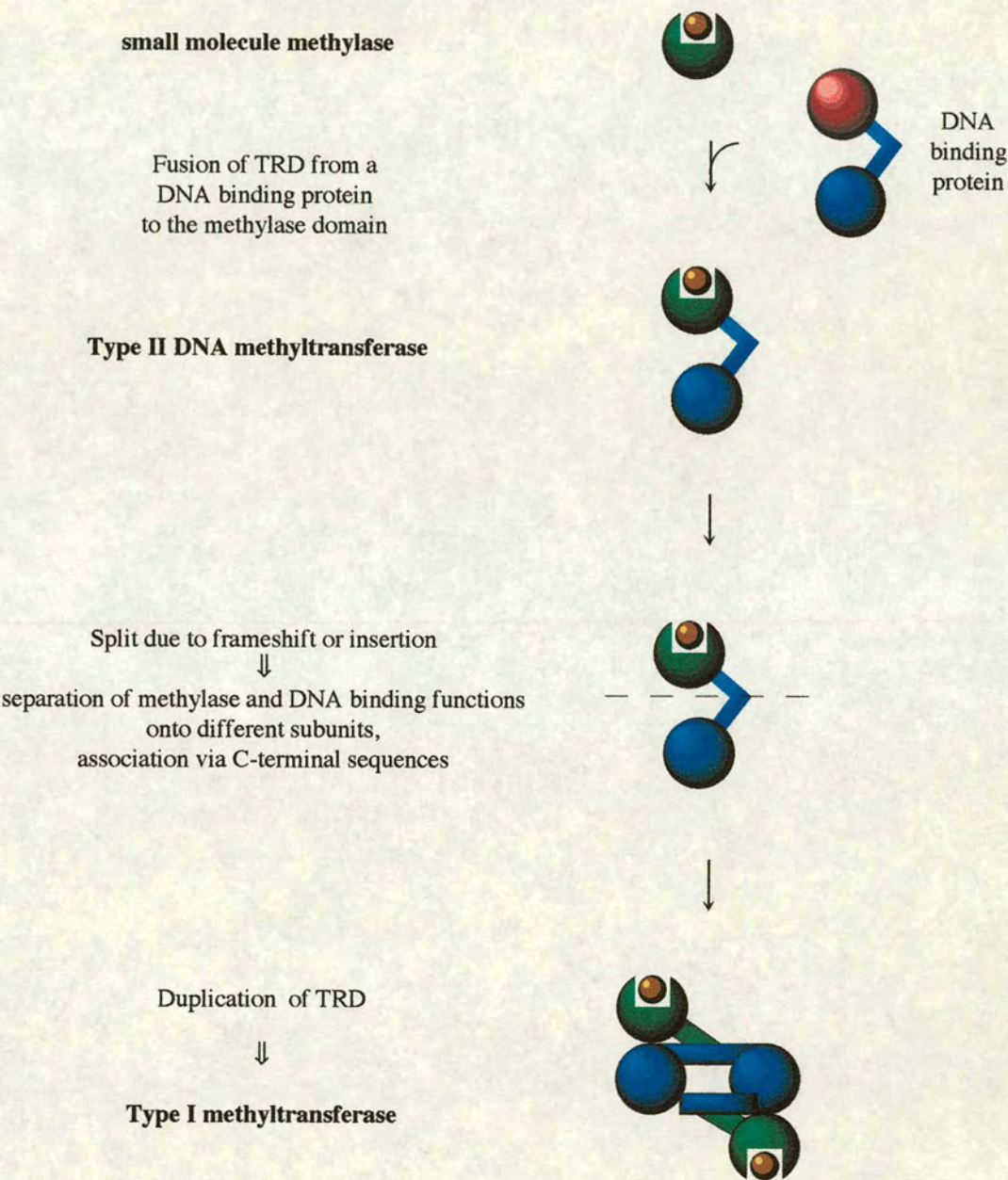


How would one envision the evolution of the complex Type IA systems? Starting with a simple methylase domain (as seen in small molecule methyltransferases) it is easy to imagine how the fusion of this domain to a DNA binding domain could produce the first simple Type II DNA methyltransferase (see fig. 6.3). In Type II systems the methyltransferase paired up with an independent restriction endonuclease, while in Type III systems the methyltransferase associated with a restriction subunit. Insertions into the methyltransferase gene of a Type III system could then put methylation and DNA binding activities onto different subunits again, with the association between the two mediated by their respective C-terminal sequences. An HsdS subunit as seen in Type I systems could result from a duplication of the gene for the DNA binding subunit, which would consequently associate with two methylation subunits. However these two methylation subunits would still function independently from each other, reminiscent of Type II methyltransferases. The subsequent addition of an N-terminal domain (possibly from an allosteric protein) to the methylation subunits results in a structure as deduced for HsdM, and opens the possibility of interactions between the two HsdM subunits. The evolution of specificity for hemi-methylated DNA however requires another step in the scheme: methylation has to be blocked for unmodified bases, but allowed if a modified base is present at the other half of the target site. It is conceivable that mutations at the start of the methylase domain (and therefore close to the AdoMet binding site) interfere with the methylation reaction. Interaction with the other HsdM subunit and the resulting repositioning of the m\* domain could slightly change the conformation of the sequence at the start of the methylase domain, thereby removing the block and allowing methylation to proceed.



**Fig. 6.3: Steps in the evolution of Type I methyltransferases**

Hypothetical model for the evolution of complex Type I systems from a simple small molecule methylase  
green: methylase domain - blue: TRD - red: unknown domain - yellow: AdoMet





**Appendix: Table of identified Type I R/M-systems**

Family	Enzyme	Target	Reference for target sequence
<b>IA</b>	<i>EcoKI</i>	AAC (N <sub>6</sub> ) GTGC	(Kan <i>et al.</i> , 1979)
	<i>EcoBI</i>	TGA (N <sub>8</sub> ) TGCT	(Lautenberger <i>et al.</i> , 1978) (Ravetch <i>et al.</i> , 1978)
	<i>EcoDI</i>	TTA (N <sub>8</sub> ) GTCY	(Nagaraja <i>et al.</i> , 1985b)
	<i>StyLTHI</i>	GAG (N <sub>6</sub> ) RTAYG	(Nagaraja <i>et al.</i> , 1985)
	<i>StySPI</i>	AAC (N <sub>6</sub> ) GTRC	(Nagaraja <i>et al.</i> , 1985)
	<i>EcoR5I</i>	new <sup>(1)</sup>	(Barcus <i>et al.</i> , 1995)
	<i>EcoR10I</i>	new <sup>(1)</sup>	(Barcus <i>et al.</i> , 1995)
	<i>EcoR11I</i>	same as <i>EcoBI</i>	(Barcus <i>et al.</i> , 1995)
	<i>EcoR12I</i>	same as <i>EcoKI</i>	(Barcus <i>et al.</i> , 1995)
	<i>EcoR13I</i>	new <sup>(1)</sup>	(Barcus <i>et al.</i> , 1995)
	<i>EcoR14I</i>	same as <i>EcoBI</i>	(Barcus <i>et al.</i> , 1995)
	<i>EcoR23I</i>	new <sup>(1)</sup>	(Barcus <i>et al.</i> , 1995)
	<i>EcoR24I</i>	same as <i>EcoKI</i>	(Barcus <i>et al.</i> , 1995)
	<i>EcoR25I</i>	same as <i>EcoR10I</i>	(Barcus <i>et al.</i> , 1995)
	<i>EcoR70I</i>	same as <i>EcoBI</i>	(Barcus <i>et al.</i> , 1995)
<b>IB</b>	<i>EcoAI</i>	GAG (N <sub>7</sub> ) GTCA	(Suri <i>et al.</i> , 1984)
	<i>EcoEI</i>	GAG (N <sub>7</sub> ) ATGC	(Cowan <i>et al.</i> , 1989)
	<i>Cfr AI</i>	GCA (N <sub>8</sub> ) GTGG	(Kannan <i>et al.</i> , 1989)
	<i>StySTI</i>	unknown	(Titheradge <i>et al.</i> , 1996)
	<i>StySKI</i>	CGAT (N <sub>7</sub> ) GTTA	(Thorpe <i>et al.</i> , 1997)
	<i>EcoR15I</i>	unknown	(Barcus <i>et al.</i> , 1995)
	<i>EcoR17I</i>	unknown	(Barcus <i>et al.</i> , 1995)
	<i>EcoR42I</i>	new <sup>(1)</sup>	(Barcus <i>et al.</i> , 1995)
<b>IC</b>	<i>EcoR124I</i>	GAA (N <sub>6</sub> ) RTCG	(Price <i>et al.</i> , 1987)
	<i>EcoDXXI</i>	TCA (N <sub>7</sub> ) RTTC	(Piekarowicz and Goguen, 1986)
	<i>EcoprrI</i>	CCA (N <sub>7</sub> ) RTGC	(Tyndall <i>et al.</i> , 1994)
	<i>Mycoplasma</i> <sup>(2)</sup>	unknown	(Dybvig and Yu, 1994)



	<i>Methanococcus</i> <sup>(2)</sup>	unknown	(Bult <i>et al.</i> , 1996)
<b>ID</b>	<i>StySBLI</i>	unknown	(Titheradge <i>et al.</i> , 1996)
	<i>EcoR9I</i>	unknown	(Barcus <i>et al.</i> , 1995)
	<i>EcoR31I</i>	unknown	(Barcus <i>et al.</i> , 1995)
	<i>EcoR65I</i>	unknown	(Barcus <i>et al.</i> , 1995)
	<i>EcoR68I</i>	unknown	(Barcus <i>et al.</i> , 1995)
	<i>KpnAI</i> <sup>(2)</sup>	unknown	(Rutebuka <i>et al.</i> , 1995)
<hr/>			
not	<i>BsuCI</i>	unknown	(Xu <i>et al.</i> , 1995)
classified	<i>Spiroplasma citri</i> <sup>(3)</sup>	unknown	(Laigret <i>et al.</i> , 1996)
	<i>KpnBI</i> <sup>(2)</sup>	unknown	(Valinluck <i>et al.</i> , 1995)
<hr/>			

- (1) Target sequences unknown but each enzyme shown to differ from other members of the same family.
- (2) Classification based on sequence similarity
- (3) partially sequenced

The underscore marks the adenine (or the thymine opposite of the adenine) that is being methylated.



# References

- Abadijjeva, A., Patel, J., Webb, M., Zinkevich, V., and Firman, K. (1993). A deletion mutant of the type IC restriction endonuclease EcoR124I expressing a novel DNA specificity. *Nucleic Acids Res.* 21, 4435-4443.
- Aggarwal, A. K. (1995). Structure and function of restriction endonucleases. *Curr. Opin. Struct. Biol.* 5, 11-19.
- Anderson, E. S., and Felix, A. (1952). Variation in Vi-Phage II of *Salmonella typhi*. *Nature* 170, 492-494.
- Anderson, J. E. (1993). Restriction endonucleases and modification methylases. *Curr. Opinion Struct. Biol.* 3, 24-30.
- Appleyard, R. K. (1954). Segregation of new lysogenic types during growth of a doubly lysogenic strain derived from *Escherichia coli* K12. *Genetics* 39, 440.
- Arber, W. (1962). Specificites biologiques de l'acide desoxyribonucleique. *Path. Microbiol.* 25, 668-681.
- Arber, W. (1965a). Host specificity of DNA produced by *Escherichia coli* V. The role of Methionine in the Production of Host Specificity. *J. Mol. Biol.* 11, 247-256.
- Arber, W. (1965b). Host-controlled modification of bacteriophage. *Ann. Rev. Microbiol.* 19, 365-378.
- Arber, W. (1970). Strain-specific restriction and modification of DNA. *Bull. Schweiz. Akad. Med. Wiss.* 25, 99-100.
- Arber, W., and Dussoix, D. (1962). Host Specificity of DNA Producted by *Escherichia Coli*: I. Host controlled modification of bacteriophage lambda. *J. Mol. Biol.* 5, 18-36.
- Arber, W., Hattman, S., and Dussoix, D. (1963). On the host-controlled modification of bacteriophage lambda. *Virology* 21, 30-35.
- Arber, W., and Morse, M. L. (1965). Host specificity of DNA produced by *Escherichia coli*. VI. Effects on bacterial conjugation. *Genetics* 51, 137-148.
- Arber, W., and Smith, J. D. (1966). Kinetics of reproduction of nucleic acids and their role in biosynthesis of proteins. *Int. Congr. Microbiol.* 5, 5.
- Argos, P. (1985). Evidence for a repeating domain in type I restriction enzymes. *EMBO J.* 4, 1351-1355.
- Athanasiadis, A., Vlassi, M., Kotsifaki, D., Tucker, P. A., Wilson, K. S., and Kokkinidis, M. (1994). Crystal structure of PvuII endonuclease reveals extensive structural homologies to EcoRV. *Struct. Biol.* 1, 469-475.
- Baker, D. J., Kan, J. L. C., and Smith, S. S. (1988). Recognition of structural perturbations in DNA by human DNA(cytosine-5)methyltransferase. *Gene* 74, 207-210.
- Baldwin, G. S., Kelly, S. M., Price, N. C., Wilson, G. W., Connolly, B. A., Artymiuk, P. J., and Hornby, D. P. (1994). Ligand-induced conformational states of the cytosine-specific DNA methyltransferase M.HgaI-2. *J. Mol. Biol.* 235, 545-553.
- Baldwin, G. S., Vipond, I. B., and Halford, S. E. (1995). Rapid reaction analysis of the catalytic cycle of the EcoRV restriction endonuclease. *Biochemistry* 34, 705-714.
- Balganesh, T. S., Reiners, L., Lauster, R., Noyer-Weidner, M., Wilke, K., and Trautner, T. A. (1987). Construction and use of chimeric SPR/phi 3T DNA methyltransferases in the definition of sequence recognizing enzyme regions. *Embo J.* 6, 3543-3549.



- Barcus, V. A., and Murray, N. E. (1995). **Barriers to recombination: restriction.** In *Population Genetics of Bacteria*, S. Baumberg, J. P. W. Young, S. R. Saunders and E. M. H. Wellington, eds. (Cambridge: Cambridge University Press), pp. 31-58.
- Barcus, V. A., Titheradge, A. J. B., and Murray, N. E. (1995). **The diversity of alleles at the hsd locus in natural populations of Escherichia coli.** *Genetics* 140, 1187-1197.
- Behrens, B., Noyer-Weidner, M., Pawlek, B., Lauster, R., Balganesch, T. S., and Trautner, T. A. (1987). **Organization of multispecific DNA methyltransferases encoded by temperate Bacillus subtilis phages.** *EMBO J.* 6, 1137-1142.
- Belogurov, A. A., Delver, E. P., and Rodzevich, O. V. (1992). **IncN plasmid pKM101 and IncII plasmid ColIb-P9 encode homologous antirestriction proteins in their leading regions.** *J. Bacteriol.* 174, 5079-5085.
- Belogurov, A. A., Yussifov, T. N., Kotova, V. U., and Zavilgelsky, G. B. (1985). **The novel gene(s) ARD of plasmid pKM101: alleviation of EcoK restriction.** *Mol. Gen. Genet.* 198, 509-513.
- Bergerat, A., and Guschlbauer, W. (1990). **The double role of methyl donor and allosteric effector of S-adenosyl-methionine for Dam methylase of E. coli.** *Nucleic Acids Res.* 18, 4369-4375.
- Bergerat, A., Guschlbauer, W., and Fazakerley, G. V. (1991). **Allosteric and catalytic binding of S-adenosylmethionine to Escherichia coli DNA adenine methyltransferase monitored by <sup>3</sup>H NMR.** *Proc. Natl. Acad. Sci. U.S.A.* 88, 6394-6397.
- Berkhout, B., and van Wamel, J. (1996). **Accurate scanning of the BssHII endonuclease in search for its DNA cleavage site.** *J. Biol. Chem.* 271, 1837-1840.
- Bertani, G., and Weigle, J. J. (1953). **Host controlled variation in bacterial viruses.** *J. Bacteriol.* 65, 113-121.
- Bickle, T. A. (1987). **DNA restriction and modification systems.** In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Megasanik, M. Schaechter and H. E. Umbarger, eds. (Washington DC: Amer. Soc. Microbiology), pp. 692-696.
- Bickle, T. A. (1993). **The ATP-dependent restriction endonucleases.** In *Nucleases*, S. M. Linn, R. S. Lloyd and R. J. Roberts, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 89-109.
- Bickle, T. A., Brack, C., and Yuan, R. (1978). **ATP-induced conformational changes in the restriction endonuclease from Escherichia coli K-12.** *Proc. Natl. Acad. Sci. U.S.A.* 75, 3099-3103.
- Birnboim, H. C., and Doly, J. (1979). **A rapid alkaline extraction procedure for screening recombinant plasmid DNA.** *Nucleic Acid Research* 7, 1513-1522.
- Bitinaite, J., Maneliene, Z., Menkevicius, S., Klimasauskas, S., Butkus, V., and Janulaitis, A. A. (1992). **Alw26I, Eco31I and Esp3I - type IIs methyltransferases modifying cytosine and adenine in complementary strands of the target DNA.** *Nucleic Acids Res.* 20, 4981-4985.
- Bocklage, H., Heeger, K., and Muller-Hill, B. (1991). **Cloning and characterization of the MboII restriction-modification system.** *Nucleic Acids Res.* 19, 1007-1013.
- Boyer, H. (1964). **Genetic control of restriction and modification in Escherichia coli.** *J. Bacteriol.* 88, 1652-1660.
- Boyer, H. W. (1971). **DNA restriction and modification mechanisms in bacteria.** *Ann. Rev. Microbiol.* 25, 153-176.
- Boyer, H. W., and Roulland-Dussoix, D. (1969). **A complementation analysis of the restriction and modification of DNA in Escherichia coli.** *J. Mol. Biol.* 41, 459-472.



- Bozic, D., Grazulis, S., Siksnys, V., and Huber, R. (1996). **Crystal structure of *Citrobacter freundii* restriction endonuclease Cfr10I at 2.15 Å resolution.** *J. Mol. Biol.* 255, 176-186.
- Branden, C., and Tooze, J. (1991). **Introduction to Protein Structure.** (New York: Garland Publishing, Inc.).
- Brooks, J. E., Nathan, P. D., Landry, D., Sznyter, L. A., Waite-Rees, P., Ives, C. L., Moran, L. S., Slatko, B. E., and Benner, J. S. (1991). **Characterization of the cloned BamHI restriction modification system: its nucleotide sequence, properties of the methylase, and expression in heterologous hosts.** *Nucleic Acids Res.* 19, 841-850.
- Bühler, R., and Yuan, R. (1978). **Characterization of a restriction enzyme from *Escherichia coli* K carrying a mutation in the modification subunit.** *J. Biol. Chem.* 253, 6756-6760.
- Bullas, L. R., Colson, C., and van Pel, A. (1976). **DNA restriction and modification systems in *Salmonella*. SQ, a new system derived by recombination between the SB system of *Salmonella typhimurium* and the SP system of *Salmonella potsdam*.** *J. Gen. Microbiol.* 95, 166-172.
- Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J. F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S. M., Weidman, J. F., Fuhrmann, J. L., Presley, E. A., Nguyen, D., Utterback, T. R., Kelley, J. M., Peterson, J. D., Sadow, P. W., Hanna, M. C., Cotton, M. D., Hurst, M. A., Roberts, K. M., Kaine, B. P., Borodovsky, M., Klenk, H. P., Fraser, C. M., Smith, H. O., Woese, C. R., and Venter, J. C. (1996). **Complete Genome Sequence of the Methanogenic Archeon *Methanococcus jannaschii*.** *Science* 273, 1058-1073.
- Burckhardt, J., Weisemann, J., Hamilton, D. L., and Yuan, R. (1981a). **Complexes formed between the restriction endonuclease EcoK and heteroduplex DNA.** *J. Mol. Biol.* 153, 425-440.
- Burckhardt, J., Weisemann, J., and Yuan, R. (1981b). **Characterization of the DNA methylase activity of the restriction enzyme from *Escherichia coli* K.** *J. Biol. Chem.* 256, 4024-4032.
- Burnet, F. M. (1946). **Virus as organism.** (Cambridge, Mass.: Harvard University Press).
- Butkus, V., Klimasauskas, S., Kersulyte, D., Vaitkevicius, D., Lebionka, A., and Janulaitis, A. A. (1985). **Investigation of restriction-modification enzymes from *M. varians* RFL19 with a new type of specificity toward modification of substrate.** *Nucleic Acids Res.* 13, 5727-5746.
- Chandrasegaran, S., and Smith, H. O. (1988). **Amino acid sequence homologies among twenty-five restriction endonucleases and methylases.** In *Structure & Expression: From Proteins to Ribosomes*, M. H. Sarma and R. H. Sarma, eds. (Guilford, New York: Adenine Press), pp. 149-156.
- Chen, A., Powell, L. M., Dryden, D. T. F., Murray, N. E., and Brown, T. (1995). **Tyrosine 27 of the specificity polypeptide of EcoKI can be UV crosslinked to a bromodeoxyuridine-substituted DNA target sequence.** *Nucleic Acids Res.* 23, 1177-1183.
- Chen, L., MacMillan, A. M., and Verdine, G. L. (1993). **Mutational separation of DNA binding from catalysis in a DNA cytosine methyltransferase.** *J. Amer. Chem. Soc.* 115, 5318-5319.
- Cheng, X., Balendiran, K., Schildkraut, I., and Anderson, J. E. (1994). **Structure of PvuII endonuclease with cognate DNA.** *EMBO J.* 13, 3927-3935.
- Cheng, X., Kumar, S., Klimasauskas, S., and Roberts, R. J. (1993a). **Crystal structure of the HhaI DNA methyltransferase.** *Cold Spring Harbor Symp. Quant. Biol.* 58, 331-338.
- Cheng, X., Kumar, S., Posfai, J., Pflugrath, J. W., and Roberts, R. J. (1993b). **Crystal structure of the HhaI DNA methyltransferase complexed with S-adenosyl-L-methionine.** *Cell* 74, 299-307.
- Chou, W.-Y., Bieber, C., and Matthews, K. S. (1989). **Tryptophan and 8-Anilino-1-naphthalenesulfonate Compete for Binding to *Trp* Repressor.** *J. Biol. Chem.* 264, 18309-18313.



- Christensen, J. R. (1961). **On the Process of Host-controlled Modification of Bacteriophage.** *Virology* 13, 40-43.
- Collins, E. B. (1956). **Host-controlled variations in bacteriophages active against lactic Streptococci.** *Virology* 2, 261-271.
- Connaughton, J. F., Kaloss, W. D., Vanek, P. G., Nardone, G. A., and Chirikjian, J. G. (1990). **The complete sequence of the Bacillus amyloliquefaciens proviral H2, BamHI methylase gene.** *Nucleic Acids Res.* 18, 4002.
- Cooper, L. P., and Dryden, D. T. F. (1994). **The domains of a type I DNA methyltransferase: interactions and role in recognition of DNA methylation.** *J. Mol. Biol.* 236, 1011-1021.
- Coulondre, C., Miller, J. H., Farabaugh, P. J., and Gilbert, W. (1987). **Molecular basis for base substitution hot spots in Escherichia coli.** *Nature* 274, 775-780.
- Cowan, G. M. (1989). **A new family of type I restriction and modification systems.** *Diss. Abstr.* 50, 441.
- Cowan, G. M., Gann, A. A. F., and Murray, N. E. (1989). **Conservation of complex DNA recognition domains between families of restriction enzymes.** *Cell* 56, 103-109.
- de la Campa, A. G., Kale, P., Springhorn, S. S., and Lacks, S. A. (1987). **Proteins encoded by the DpnII restriction gene cassette. Two methylases and an endonuclease.** *J. Mol. Biol.* 196, 457-469.
- Degtyarev, S. K., Rechkunova, N. I., Zernov, Y. P., Dedkov, V. S., Chizikov, V. E., Van Calligan, M., Williams, R., and Murray, E. (1993). **Bsp24I, a new unusual restriction endonuclease.** *Gene* 131, 93-95.
- Drexler, H., and Christensen, J. R. (1961). **Genetic crosses between restricted and unrestricted phage T1 in lysogenic and nonlysogenic hosts.** *Virology* 13, 31-39.
- Dryden, D. T. F., Cooper, L. P., and Murray, N. E. (1993). **Purification and characterization of the methyltransferase from the Type I restriction and modification system of Escherichia coli K12.** *J. Biol. Chem.* 268, 13228-13236.
- Dryden, D. T. F., Cooper, L. P., Thorpe, P. H., and Byron, O. (1997). **The in vitro assembly of the EcoKI type I DNA restriction/modification enzyme and its in vivo implications.** *Biochemistry* 36, 1065-1076.
- Dryden, D. T. F., Sturrock, S. S., and Winter, M. (1995). **Structural modelling of a type I DNA methyltransferase.** *Nature Struct. Biol.* 2, 632-635.
- Dubey, A. K., and Roberts, R. J. (1992). **Sequence-specific DNA binding by the MspI DNA methyltransferase.** *Nucleic Acids Res.* 20, 3167-3173.
- Duncan, B. K., and Miller, J. H. (1980). **Mutagenic deamination of cytosine residues in DNA.** *Nature* 287, 560-561.
- Dunn, D. B., and Smith, J. D. (1955). **Occurrence of a new base in the deoxyribonucleic acid of a strain of Bacterium coli.** *Nature* 175, 336-337.
- Dussoix, D., and Arber, W. (1962). **Host Specificity of DNA Produced by Escherichia Coli: II. Control over acceptance of DNA from infecting phage lambda.** *J. Mol. Biol.* 5, 37-49.
- Dybvig, K., and Yu, H. (1994). **Regulation of a restriction and modification system via DNA inversion in Mycoplasma pulmonis.** *Mol. Microbiol.* 12, 547-560.
- Ehbrecht, H.-J., Pingoud, A., Urbanke, C., Maass, G., and Gualerzi, C. (1985). **Linear diffusion of restriction endonucleases on DNA.** *J. Biol. Chem.* 260, 6160-6166.
- Ehrlich, M., Gama-Sosa, M. A., Carreira, L. H., Ljungdahl, L. G., Kuo, K. C., and Gehrke, C. W. (1985). **DNA methylation in thermophilic bacteria: N4-methylcytosine, 5-methylcytosine, and N6-methyladenine.** *Nucleic Acids Res.* 13, 1399-1412.



- Ehrlich, M., Wilson, G. G., Kenneth, C. K., and Gehrke, C. W. (1987). **N4-Methylcytosine as a minor base in bacterial DNA.** *J. Bacteriol.* **169**, 939-943.
- Erlanson, D. A., Chen, L., and Verdine, G. L. (1993). **DNA methylation through a locally upaired intermediate.** *J. Am. Chem. Soc.* **115**, 12583-12584.
- Eskin, B., and Linn, S. (1972). **The deoxyribonucleic acid modification and restriction enzymes of *Escherichia coli* B. III. Studies of the restriction adenosine triphosphatase.** *J. Biol. Chem.* **247**, 6192-6196.
- Findlay, G. M. (1939). **Variation in Viruses.** In *Handbuch der Virusforschung*, D. a. Hallauer, ed. (Vienna: Springer).
- Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., and al., e. (1995). **Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.** *Science* **269**, 496-512.
- Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., and al., e. (1995). **The minimal gene complement of *Mycoplasma genitalium*.** *Science* **270**, 397-403.
- Frederick, C. A., Grable, J., Melia, M., Samudzi, C., Jen-Jacobson, L., Wang, B.-C., Greene, P., Boyer, H. W., and Rosenberg, J. M. (1984). **Kinked DNA in crystalline complex with *EcoRI* endonuclease.** *Nature* **309**, 327-331.
- Fried, M. G. (1989). **Measurement of protein-DNA interaction parameters by electrophoresis mobility shift assay.** *Electrophoresis* **10**, 366-376.
- Friedman, S., and Ansari, N. (1992). **Binding of the *EcoRII* methyltransferase to 5-fluorocytosine-containing DNA. Isolation of a bound peptide.** *Nucleic Acids Res.* **20**, 3241-3248.
- Fuller-Pace, F. V. (1994). **RNA helicases: Modulators of RNA structure.** *Trends Cell. Biol.* **4**, 271-274.
- Fuller-Pace, F. V., Bullas, L. R., Delius, H., and Murray, N. E. (1984). **Genetic recombination can generate altered restriction specificity.** *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6095-6099.
- Fuller-Pace, F. V., and Murray, N. E. (1986). **Two DNA recognition domains of the specificity polypeptides of a family of type I restriction enzymes.** *Proc. Natl. Acad. Sci. U.S.A.* **83**, 9368-9372.
- Fürste, J. P., Pansegrau, W., Frank, R., Blöcker, H., Scholz, P., Bagdasarian, M., and Lanka, E. (1986). **Molecular cloning of the plasmid RP4 primase region in a multi-host-range *tacP* expression vector.** *Gene* **48**, 119-131.
- Gann, A. A. F., Campbell, A. J. B., Collins, J. F., Coulson, A. F. W., and Murray, N. E. (1987). **Reassortment of DNA recognition domains and the evolution of new specificities.** *Mol. Microbiol.* **1**, 13-22.
- Glover, S. W. (1970). **Functional analysis of host-specificity mutants in *Escherichia coli*.** *Genet. Res.* **15**, 237-250.
- Glover, S. W., and Colson, C. (1969). **Genetics of host-controlled restriction and modification in *Escherichia coli*.** *Genet. Res.* **13**, 227-240.
- Glover, S. W., Firman, K., Watson, G., Price, C., and Donaldson, S. (1983). **The alternate expression of two restriction and modification systems.** *Mol. Gen. Genet.* **190**, 65-69.
- Glover, S. W., Schell, J., Symonds, N., and Stacey, K. A. (1963). **The control of host-induced modification by phage P1.** *Genet. Res.* **4**, 480-482.
- Gold, M., and Hurwitz, J. (1963). **The Enzymatic Methylation of the Nucleic Acids.** In *Cold Spring harbor symposia on quantitative biology* (Cold Spring Harbor, Long Island, N. Y.: Long Island Biological Association), pp. 149-156.



- Gold, M., Hurwitz, J., and Anders, M. (1963a). The Enzymatic Methylation of RNA and DNA. I. Biochem. Biophys. Res. Comm. 11, 107-114.
- Gold, M., Hurwitz, J., and Anders, M. (1963b). The Enzymatic Methylation of RNA and DNA. II. On the Species Specificity of the Methylation Enzymes. PNAS 50, 164-169.
- Gong, W., O'Gara, M., Blumenthal, R. M., and Cheng, X. (1997). Structure of PvuII DNA-(cytosine N4) methyltransferase, an example of domain permutation and protein fold assignment. Nucleic Acids Res. 25, 2702-2715.
- Gopal, J., Yebra, M. J., and Bhagwat, A. S. (1994). DsaV methyltransferase and its isoschizomers contain a conserved segment that is similar to the segment in HhaI methyltransferase that is in contact with DNA bases. Nucleic Acids Res. 22, 4482-4488.
- Gorbalenya, A. E., and Koonin, E. V. (1991). Endonuclease (R) subunits of type-I and type-III restriction-modification enzymes contain a helicase-like domain. FEBS Lett. 291, 277-281.
- Gough, J. A., and Murray, N. E. (1983). Sequence diversity among related genes for recognition of specific targets in DNA molecules. J. Mol. Biol. 166, 1-19.
- Gough, M., and Lederberg, S. (1966). Methylated bases in the host-modified deoxyribonucleic acid of Escherichia coli and bacteriophage lambda. J. Bacteriol. 91, 1460-1468.
- Grasby, J. A., and Connolly, B. A. (1992). Stereochemical outcome of the hydrolysis reaction catalyzed by the EcoRV restriction endonuclease. Biochemistry 31, 7855-7861.
- Grasso, R. J., and Paigen, K. (1968a). Loss of host-controlled restriction of lambda bacteriophage in Escherichia coli following methionine deprivation. J. Virol. 2, 1368-1373.
- Grasso, R. J., and Paigen, K. (1968b). The effect of amino acids on host-controlled restriction of lambda phage. Virology 36, 1-8.
- Gubler, M., and Bickle, T. A. (1991). Increased protein flexibility leads to promiscuous protein-DNA interactions in type IC restriction-modification systems. EMBO J. 10, 951-957.
- Gubler, M., Bragaglia, D., Meyer, J., Piekarowicz, A., and Bickle, T. A. (1992). Recombination of constant and variable modules alters DNA sequence recognition by type IC restriction-modification enzymes. Embo J. 11, 233-240.
- Gunthert, U., Jentsch, S., and Freund, M. (1981). Restriction and modification in Bacillus subtilis: Two DNA Methyltransferases with BsuRI specificity. II. Catalytic properties, substrate specificity, and mode of action. J. Biol. Chem. 256, 9346-9351.
- Haberman, A. (1974). The bacteriophage P1 restriction endonuclease. J. Mol. Biol. 89, 545-563.
- Hadi, S. M., Bickle, T. A., and Yuan, R. (1975). The role of S-adenosylmethionine in the cleavage of deoxyribonucleic acid by the restriction endonuclease from Escherichia coli K. J. Biol. Chem. 250, 4159-4164.
- Hanck, T., Schmidt, S., and Fritz, H. J. (1993). Sequence-specific and mechanism-based crosslinking of Dcm DNA cytosine-C5 methyltransferase of E.coli K-12 to synthetic oligonucleotides containing 5-fluoro-2'-deoxycytidine. Nucleic Acids Res. 21, 303-309.
- Harris, D. A. (1987). Spectrophotometric assays. In Spectrophotometry and Spectrofluorimetry., D. A. Harris and C. L. Bushford, eds. (Oxford: IRL Press), pp. 49-90.
- Hendrickson, W., and Schleif, R. F. (1984). Regulation of the Escherichia coli L-arabinose operon studied by the gel electrophoresis binding assay. J. Mol. Biol. 174, 611-628.



- Highlander, S. K., and Garza, O. (1996). The restriction-modification system of *Pasteurella haemolytica* is a member of a new family of type I enzymes. *Gene* 178, 89-96.
- Hill, C., Miller, L. A., and Klaenhammer, T. R. (1991). In vivo genetic exchange of a functional domain from a type II A methylase between lactococcal plasmid pTR2030 and a virulent bacteriophage. *J. Bacteriol.* 173, 4363-4370.
- Himmelreich, R., Hilbert, H., Plagens, H., Pirkel, E., Li, B. C., and Herrmann, R. (1996). Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res.* 24, 4420-4449.
- Hirsch, J. A., Wah, D. A., Dorner, L. F., Schildkraut, I., and Aggarwal, A. K. (1997). Crystallization and preliminary X-ray analysis of restriction endonuclease FokI bound to DNA. *FEBS Lett.* 403, 136-138.
- Ho, D. K., Wu, J. C., Santi, D. V., and Floss, H. G. (1991). Stereochemical studies of the C-methylation of deoxycytidine catalyzed by HhaI methylase and the N-methylation of deoxyadenosine catalyzed by EcoRI methylase. *Arch. Biochem. Biophys.* 284, 264-269.
- Holmes, D. S., and Quigley, M. (1981). A rapid boiling method for the preparation of bacterial plasmids. *Anal. Biochem.* 114, 193.
- Huang, B., Schaeffer, C. J., Li, Q., and Tsai, M. D. (1996). Splase: A new class IIS zinc-finger restriction endonuclease with specificity for SpI binding sites. *J. Protein Chem.* 15, 481-489.
- Iida, S., Meyer, J., Bachi, B., Stalhammar-Carlemalm, M., Schrickel, S., Bickle, T. A., and Arber, W. (1983). DNA restriction--modification genes of phage P1 and plasmid p15B. Structure and in vitro transcription. *J. Mol. Biol.* 165, 1-18.
- Ish-Horowicz, D., and Burke, J. F. (1981). Rapid and efficient cosmid cloning. *Nucleic Acid Research* 9, 2989.
- Janscak, P., Abadjieva, A., and Firman, K. (1996). The type I restriction endonuclease R.EcoR124I: over-production and biochemical properties. *J. Mol. Biol.* 257, 977-991.
- Janulaitis, A., Petrusyte, M., Maneliene, Z., Klimasauskas, S., and Butkus, V. (1992a). Purification and properties of the Eco57I restriction endonuclease and methylase--prototypes of a new class (type IV). *Nucleic Acids Res.* 20, 6043-6049.
- Janulaitis, A., Vaisvila, R., Timinskas, A., Klimasauskas, S., and Butkus, V. (1992b). Cloning and sequence analysis of the genes coding for Eco57I type IV restriction-modification enzymes. *Nucleic Acids Res.* 20, 6051-6056.
- Janulaitis, A. A., Kazlauskienė, R., Lazareviciute, L., Gilvonauskaite, R., Steponaviciene, D., Jagelavicius, M., Petrusyte, M., Bitinaite, J., Vezeviciute, Z., Kiuduliene, E., and Butkus, V. (1988). Taxonomic specificity of restriction-modification enzymes. *Gene* 74, 229-232.
- Janulaitis, A. A., Klimasauskas, S., Petrusyte, M., and Butkus, V. (1983). Cytosine modification in DNA by BcnI methylase yields N4-methylcytosine. *FEBS Lett.* 161, 131-134.
- Jeltsch, A., Alves, J., Maass, G., and Pingoud, A. (1992). On the Catalytic Mechanism of EcoRI and EcoRV. *FEBS Letters* 304, 4-8.
- Jeltsch, A., Alves, J., Wolfes, H., Maass, G., and Pingoud, A. (1994). Pausing of the restriction endonuclease EcoRI during linear diffusion on DNA. *Biochemistry* 33, 10215-10219.
- Jeltsch, A., Kroger, M., and Pingoud, A. (1995b). Evidence for an evolutionary relationship among type-II restriction endonucleases. *Gene* 160, 7-16.
- Jeltsch, A., Pleckaityte, M., Selent, U., Wolfes, H., Siksnys, V., and Pingoud, A. (1995b). Evidence for substrate-assisted catalysis in the DNA cleavage of several restriction endonucleases. *Gene* 157, 157-162.



- Jo, K., and Topal, M. D. (1995). **DNA topoisomerase and recombinase activities in NaeI restriction endonuclease.** *Science* 267, 1817-1820.
- Kan, N. C., Lautenberger, J. A., Edgell, M. H., and Hutchison, C. A. I. (1979). **The nucleotide sequence recognized by the Escherichia coli K12 restriction and modification enzymes.** *J. Mol. Biol.* 130, 191-209.
- Kannan, P., Cowan, G. M., Daniel, A. S., Gann, A. A. F., and Murray, N. E. (1989). **Conservation of organization in the specificity polypeptides of two families of Type I restriction enzymes.** *J. Mol. Biol.* 209, 335-344.
- Kauc, L., and Piekarowicz, A. (1978). **Purification and properties of a new restriction endonuclease from Haemophilus influenzae Rf.** *Eur. J. Biochem.* 92, 417-426.
- Kelleher, J. E., Daniel, A. S., and Murray, N. E. (1991). **Mutations that confer *de novo* activity upon a maintenance methyltransferase.** *J. Mol. Biol.* 221, 431-440.
- Kellenberger, G., Symonds, N., and Arber, W. (1966). **Host specificity of DNA produced by Escherichia coli. 8. Its acquisition by phage lambda and its persistence through consecutive growth cycles.** *Z. Vererbungsl.* 98, 247-256.
- Kelly, T. J. J., and Smith, H. O. (1970). **A restriction enzyme from Hemophilus influenzae II. Base sequence of the recognition site.** *J. Mol. Biol.* 51, 393-409.
- Kim, S. C., Posfai, G., and Szybalski, W. (1991). **A novel gene-fusing vector: construction of a 5'-GGmCC-specific chimeric methyltransferase, M.BspRI/M.BsuRI.** *Gene* 100, 45-50.
- Kim, Y., and Chandrasegaran, S. (1994). **Chimeric restriction endonuclease.** *Proc. Natl. Acad. Sci. U.S.A.* 91, 883-887.
- Kim, Y., Grable, J. C., Love, R., Greene, P. J., and Rosenberg, J. M. (1990). **Refinement of EcoRI endonuclease crystal structure: A revised protein chain tracing.** *Science* 249, 1307-1309.
- Kim, Y.-G., Cha, J., and Chandrasegaran, S. (1996). **Hybrid restriction enzymes: Zinc finger fusions to FokI cleavage domain.** *Proc. Natl. Acad. Sci. U.S.A.* 93, 1156-1160.
- Kim, Y.-G., Li, L., and Chandrasegaran, S. (1994). **Insertion and deletion mutants of FokI restriction endonuclease.** *J. Biol. Chem.* 269, 31978-31982.
- King, G., and Murray, N. E. (1995). **Restriction alleviation and modification enhancement by the Rac prophage of Escherichia coli K-12.** *Mol. Microbiol.* 16, 769-777.
- Kita, K., Kotani, H., Hiraoka, N., Nakamura, T., and Yonaha, K. (1989). **Overproduction and crystallization of FokI restriction endonuclease.** *Nucleic Acids Res.* 17, 8741.
- Kita, K., Suisha, M., Kotani, H., Yanase, H., and Kato, N. (1992). **Cloning and sequence analysis of the StsI restriction-modification gene: presence of homology to FokI restriction-modification enzymes.** *Nucleic Acids Res.* 20, 4167-4172.
- Klein, A. (1965). **Mechanisms of host-controlled modification of phage T1.** *Z. Vererbungsl.* 96, 346-363.
- Klein, A., and Sauerbier, W. (1965). **Role of methylation in host controlled modification of phage T1.** *Biochem. Biophys. Res. Commun.* 18, 440-445.
- Klimasauskas, S., Kumar, S., Roberts, R. J., and Cheng, X. (1994). **HhaI methyltransferase flips its target base out of the DNA helix.** *Cell* 76, 357-369.
- Klimasauskas, S., Nelson, J. L., and Roberts, R. J. (1991). **The sequence specificity domain of cytosine-C5 methylases.** *Nucleic Acids Res.* 19, 6183-6190.
- Klimasauskas, S., and Roberts, R. J. (1995). **M.HhaI binds tightly to substrates containing mismatches at the target base.** *Nucleic Acids Res.* 23, 1388-1395.



- Klimasauskas, S., Timinskas, A., Menkevicius, S., Butkiene, D., Butkus, V., and Janulaitis, A. A. (1989). **Sequence motifs characteristic of DNA[cytosine-N4]methylases: similarity to adenine and cytosine-C5 DNA-methylases.** *Nucleic Acids Res.* 17, 9823-9832.
- Klimasauskas, S., Timinskas, A., Menkevicius, S., Butkiene, D., Butkus, V., and Janulaitis, A. A. (1990). **Sequence motifs characteristic of DNA-cytosine-N4-methyltransferases: The two domains of global similarity within DNA-methylases.** *Eksp. Biol.* 0, 4-12.
- Kneale, G. G. (1994). **A symmetrical model for the domain structure of Type I DNA methyltransferases.** *J. Mol. Biol.* 243, 1-5.
- Kong, H., Morgan, R. D., Maunus, R. E., and Schildkraut, I. (1993). **A unique restriction endonuclease, BcgI, from Bacillus coagulans.** *Nucleic Acids Res.* 21, 987-991.
- Kong, H., Roemer, S. E., Waite-Rees, P. A., Benner, J. S., Wilson, G. G., and Nwankwo, D. O. (1994). **Characterization of BcgI, a new kind of restriction-modification system.** *J. Biol. Chem.* 269, 683-690.
- Kostrewa, D., and Winkler, F. K. (1995). **Mg<sup>2+</sup> binding to the active site of EcoRV endonuclease: A crystallographic study of complexes with substrate and product DNA at 2 Angstrom resolution.** *Biochemistry* 34, 683-696.
- Krüger, D. H., and Bickle, T. A. (1983). **Bacteriophage Survival: Multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts.** *Microbiol. Rev.* 47, 345-360.
- Krüger, D. H., Bickle, T. A., Reuter, M., Pein, C.-D., and Schroeder, C. (1990). **DNA methylation and restriction processes in Escherichia coli: Insights by use of bacterial viruses T3 and T7.** In *Nucleic Acid Methylation*, G. A. Clawson, D. B. Willis, A. Weissbach and P. A. Jones, eds. (New York: Alan R. Liss, Inc), pp. 113-124.
- Krüger, D. H., Hansen, S., and Schroeder, C. (1980a). **Host-dependent modification of bacterial virus T3 affecting its adsorption ability.** *Virology* 102, 444-446.
- Krüger, D. H., Hansen, S., and Schroeder, C. (1980b). **Virus adaptation to host cells: the non-classical modification of phage T3.** *Z. Allg. Mikrobiol.* 20, 495-502.
- Krüger, D. H., and Schroeder, C. (1981). **Bacteriophage T3 and bacteriophage T7 virus-host cell interactions.** *Microbiol. Rev.* 45, 9-51.
- Krüger, D. H., Schroeder, C., Santibanez-Koref, M., and Reuter, M. (1989). **Avoidance of DNA methylation: A virus-encoded methylase inhibitor and evidence for counterselection of methylase recognition sites in viral genomes.** *Cell Biophys.* 15, 87-95.
- Kumar, S., Cheng, X., Klimasauskas, S., Sha, M., Posfai, J., Roberts, R. J., and Wilson, G. G. (1994). **The DNA (cytosine-5) methyltransferases.** *Nucleic Acids Res.* 22, 1-10.
- Kumar, S., Cheng, X., Pflugrath, J. W., and Roberts, R. J. (1992). **Purification, crystallization, and preliminary X-ray diffraction analysis of an M.HhaI-AdoMet complex.** *Biochemistry* 31, 8648-8653.
- Kusiak, M., Price, C., Rice, D., and Hornby, D. P. (1992). **The HsdS polypeptide of the Type IC restriction enzyme EcoR124 is a sequence-specific DNA-binding protein.** *Mol. Microbiol.* 6, 3251-3256.
- Labahn, J., Granzin, J., Schluckebier, G., Robinson, D. P., Jack, W. E., Schildkraut, I., and Saenger, W. (1994). **Three-dimensional structure of the adenine-specific DNA methyltransferase M.TaqI in complex with the cofactor S-adenosylmethionine.** *Proc. Natl. Acad. Sci. U.S.A.* 91, 10957-10961.
- Laemmli, U. K. (1970). **Cleavage of structural proteins during the assembly of the head of bacteriophage T4.** *Nature* 227, 680-685.



- Laigret, F., Gaurivaud, P., and Bove, J.-M. (1996). The unique organization of the *rpoB* region of *Spiroplasma citri*: a restriction and modification system gene is adjacent to *rpoB*. *Gene* 171, 95-98.
- Landry, D., Looney, M. C., Feehery, G. R., Slatko, B. E., Jack, W. E., Schildkraut, I., and Wilson, G. G. (1989). *M.FokI* methylates adenine in both strands of its asymmetric recognition sequence. *Gene* 77, 1-10.
- Lauster, R. (1989). Evolution of type II DNA methyltransferases: A gene duplication model. *J. Mol. Biol.* 206, 313-321.
- Lauster, R., Trautner, T. A., and Noyer-Weidner, M. (1989). Cytosine-specific type II DNA methyltransferases: A conserved enzyme core with variable target-recognizing domains. *J. Mol. Biol.* 206, 305-312.
- Lautenberger, J. A., Kan, N. C., Lackey, D., Linn, S., Edgell, M. H., and Hutchison, C. A. I. (1978). Recognition site of *Escherichia coli* B restriction enzyme on *PhiXsB1* and simian virus 40 DNAs: An interrupted sequence. *Proc. Natl. Acad. Sci. U.S.A.* 75, 2271-2275.
- Lautenberger, J. A., and Linn, S. (1972). The deoxyribonucleic acid modification and restriction enzymes of *Escherichia coli* B. I. Purification, subunit structure, and catalytic properties of the modification methylase. *J. Biol. Chem.* 247, 6176-6182.
- Leatherbarrow, R. J. (1992). *GraFit*. Version 3.0, Erithacus Software Ltd., Staines, U.K.
- Lederberg, S. (1957). Suppression of the Multiplication of Heterologous Bacteriophages in Lysogenic Bacteria. *Virology* 3, 496-513.
- Ledinko, N. (1964). Occurrence of 5-Methyldeoxycytidylate in the DNA of Phage Lambda. *J. Mol. Biol.* 9, 834-835.
- Levy, W. P., and Welker, N. E. (1981). Deoxyribonucleic acid modification methylase from *Bacillus stearothermophilus*. *Biochemistry* 20, 1120-1127.
- Li, L. (1994). Studies on *FokI* (a type IIS) restriction endonuclease. *Diss. Abstr.* 55, 877B-878B.
- Li, L., Wu, L. P., and Chandrasegaran, S. (1992). Functional domains in *FokI* restriction endonuclease. *Proc. Natl. Acad. Sci. U.S.A.* 89, 4275-4279.
- Li, L., Wu, L. P., Clarke, R., and Chandrasegaran, S. (1993). C-terminal deletion mutants of the *FokI* restriction endonuclease. *Gene* 133, 79-84.
- Linder, P., Lasko, P. F., Ashburner, M., Leroy, P., Nielsen, P. J., Nishi, K., Schnier, J., and Slonimski, P. P. (1989). Birth of the D-E-A-D box. *Nature* 337, 121-122.
- Linn, S., and Arber, W. (1968). Host specificity of DNA produced by *Escherichia coli*, X. In vitro restriction of phage fd replicative form. *Proc. Natl. Acad. Sci. U.S.A.* 59, 1300-1306.
- Loenen, W. A. M., and Murray, N. E. (1986). Modification enhancement by the restriction alleviation protein (Ral) of bacteriophage lambda. *J. Mol. Biol.* 190, 11-22.
- Looney, M. C., Moran, L. S., Jack, W. E., Feehery, G. R., Benner, J. S., Slatko, B. E., and Wilson, G. G. (1989). Nucleotide sequence of the *FokI* restriction-modification system: separate strand-specificity domains in the methyltransferase. *Gene* 80, 193-208.
- Lubys, A., Lubiene, J., Kulakauskas, S., Stankevicius, K., Timinskas, A., and Janulaitis, A. (1996). Cloning and analysis of the genes encoding the type IIS restriction-modification system *HphI* from *Haemophilus parahaemolyticus*. *Nucleic Acids Res.* 24, 2760-2766.
- Luria, S. E. (1945). Mutations of bacterial viruses affecting their host range. *Genetics* 30, 84-99.
- Luria, S. E. (1953). Host-induced modifications of viruses. *Cold Spring Harbor Symp. Quant. Biol.* 18, 237-244.



- Luria, S. E., and Human, M. L. (1952). A nonhereditary, host-induced variation of bacterial viruses. *J. Bacteriol.* 64, 557-569.
- Maass, G., Alves, J., Fliess, A., Pingoud, A., and Wolfes, H. (1986). Recognition of DNA-sequences by restriction endonucleases. *Hoppe-Seyler's Z. Physiol. Chem.* 367, 47.
- MacWilliams, M. P., and Bickle, T. A. (1996). Generation of new DNA binding specificity by truncation of the type IC EcoDXXI hsdS gene. *Embo J.* 15, 4775-4783.
- Maegley, K. A., Gonzalez, L., Smith, D. W., and Reich, N. O. (1992). Cofactor and DNA interactions in EcoRI DNA methyltransferase. *J. Biol. Chem.* 267, 18527-18532.
- Malone, T., Blumenthal, R. M., and Cheng, X. (1995). Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyl-transferases, and suggests a catalytic mechanism for these enzymes. *J. Mol. Biol.* 253, 618-632.
- Mamelak, L., and Boyer, H. W. (1970). Genetic control of the secondary modification of deoxyribonucleic acid in *Escherichia coli*. *J. Bacteriol.* 104, 57-62.
- Mann, M. B., and Smith, H. O. (1979). Specificity of DNA Methylases from *Haemophilus* sp. In *Proceedings of the Conference on Transmethylation*, E. Usdin, R. T. Borchardt and C. R. Creveling, eds. (New York: Elsevier/North Holland), pp. 483-492.
- Marinus, M. G. (1987). DNA methylation in *Escherichia coli*. *Annu. Rev. Genet.* 21, 113-131.
- Marinus, M. G., and Morris, N. R. (1973). Isolation of deoxyribonucleic acid methylase mutants of *Escherichia coli* K-12. *J. Bacteriol.* 114, 1143-1150.
- McClarín, J. A., Frederick, C. A., Wang, B. C., Greene, P., Boyer, H. W., Grable, J., and Rosenberg, J. M. (1986). Structure of the DNA-EcoRI endonuclease recognition complex at 3 angstrom resolution. *Science* 234, 1526-1541.
- Meisel, A., Bickle, T. A., Kruger, D. H., and Schroeder, C. (1992). Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. *Nature* 355, 467-469.
- Meisel, A., Mackeldanz, P., Bickle, T. A., Kruger, D. H., and Schroeder, C. (1995). Type III restriction endonucleases translocate DNA in a reaction driven by recognition site-specific ATP hydrolysis. *EMBO J.* 14, 2958-2966.
- Meister, J., MacWilliams, M., Hubner, P., Jutte, H., Skrzypek, E., Piekarowicz, A., and Bickle, T. A. (1993). Macroevolution by transposition: drastic modification of DNA recognition by a type I restriction enzyme following Tn5 transposition. *Embo J.* 12, 4585-4591.
- Merkel, R., Kroeger, M., Rice, P., and Fritz, H.-J. (1992). Statistical evaluation and biological interpretation of non-random abundance in the *E. coli* K-12 genome of tetra- and pentanucleotide sequences related to VSP DNA mismatch repair. *Nucleic Acids Res.* 20, 1657-1662.
- Mernagh, D. R., and Kneale, G. G. (1996). High resolution footprinting of a type I methyltransferase reveals a large structural distortion within the DNA recognition site. *Nucleic Acids Res.* 24, 4853-4858.
- Meselson, M., and Yuan, R. (1968). DNA restriction enzyme from *E. coli*. *Nature* 217, 1110-1114.
- Mi, S., and Roberts, R. J. (1992). How MspI and HpaII decide which base to methylate. *Nucleic Acids Res.* 20, 4811-4816.
- Mihalyi, E. (1970). Numeric values of the absorbance of the aromatic amino acids in acid, neutral, and alkaline solutions. In *Handbook of Biochemistry*, H. A. Sober, ed. (Ohio: CRC Press), pp. B75-B77.
- Monod, J., Wyman, J., and Changeux, J.-P. (1965). On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* 12, 88-118.



- Murray, N. E., Batten, P. L., and Murray, K. (1973). **Restriction of bacteriophage lambda by *Escherichia coli* K.** *J. Mol. Biol.* *81*, 395-407.
- Murray, N. E., Daniel, A. S., Cowan, G. M., and Sharp, P. M. (1993). **Conservation of motifs within the unusually variable polypeptide sequences of type I restriction and modification enzymes.** *Mol. Microbiol.* *9*, 133-143.
- Nagaraja, V., Shepherd, J. C., Pripfl, T., and Bickle, T. A. (1985). **Two type I restriction enzymes from *Salmonella* species. Purification and DNA recognition sequences.** *J. Mol. Biol.* *182*, 579-587.
- Nagaraja, V., Shepherd, J. C. W., and Bickle, T. A. (1985a). **A hybrid recognition sequence in a recombinant restriction enzyme and the evolution of DNA sequence specificity.** *Nature* *316*, 371-372.
- Nagaraja, V., Stieger, M., Nager, C., Hadi, S. M., and Bickle, T. A. (1985b). **The nucleotide sequence recognized by the *Escherichia coli* D type I restriction and modification enzyme.** *Nucleic Acids Res.* *13*, 389-399.
- Newman, M., Strzelecka, T., Dorner, L. F., Schildkraut, I., and Aggarwal, A. K. (1994). **Structure of restriction endonuclease BamHI phased at 1.95 Å resolution by MAD analysis.** *Structure* *2*, 439-452.
- Newman, M., Strzelecka, T., Dorner, L. F., Schildkraut, I., and Aggarwal, A. K. (1995). **Structure of BamHI endonuclease bound to DNA: partial folding and unfolding on DNA binding.** *Science* *269*, 656-663.
- Noyer-Weidner, M., Walter, J., Terschuren, P. A., Chai, S., and Trautner, T. A. (1994). **M.phi 3TII: a new monospecific DNA (cytosine-C5) methyltransferase with pronounced amino acid sequence similarity to a family of adenine-N6-DNA- methyltransferases.** *Nucleic Acids Res.* *22*, 4066-4072.
- O'Gara, M., McCloy, K., Malone, T., and Cheng, X. (1995). **Structure-based sequence alignment of three AdoMet-dependent DNA methyltransferases.** *Gene* *157*, 135-138.
- O'Gara, M., S., K., R.J., R., and X., C. (1996). **Enzymatic C5-cytosine methylation of DNA: Mechanistic implications of new crystal structures for HhaI methyltransferase-DNA-AdoHcy complexes.** *J. Mol. Biol.* *261*, 634-645.
- Oda, K., and Marmur, J. (1966). **Purification and properties of deoxyribonucleic acid methylase from *Bacillus subtilis*.** *Biochemistry* *5*, 761-773.
- Patel, D. J. (1994). **A molecular handshake.** *Nature* *367*, 688-690.
- Petrusyte, M., Bitinaite, J., Menkevicius, S., Klimasauskas, S., Butkus, V., and Janulaitis, A. A. (1988). **Restriction endonucleases of a new type.** *Gene* *74*, 89-91.
- Phillips, S. E. V. (1994). **Induced flip.** *Nature Struct. Biol.* *1*, 76-77.
- Piekarowicz, A., and Goguen, J. D. (1986). **The DNA sequence recognized by the EcoDXXI restriction endonuclease.** *Eur. J. Biochem.* *154*, 295-298.
- Piekarowicz, A., Goguen, J. D., and Skrzypek, E. (1985). **The EcoDXXI restriction and modification system of *Escherichia coli* ET7. Purification, subunit structure and properties of the restriction endonuclease.** *Eur. J. Biochem.* *152*, 387-393.
- Piekarowicz, A., Yuan, R., and Stein, D. C. (1988). **Purification and characterization of DNA methyltransferases from *Neisseria gonorrhoeae*.** *Nucleic Acids Res.* *16*, 5957-5972.
- Pingoud, A., Jeltsch, A., Lanio, T., Schulze, C., Selent, U., Stahl, F., Wende, W., and Wenz, C. (1995). **Mechanism of DNA recognition and cleavage by restriction enzymes.** *Biol. Chem. Hoppe-Seyler* *376*, S138.



- Pittard, J. (1964). **Effect of phage-controlled restriction on genetic links in bacterial crosses.** *J. Bacteriol.* 87, 1256-1257.
- Pogolotti, A. L., Ono, A., Subramaniam, R., and Santi, D. V. (1988). **On the mechanism of DNA-adenine methylase.** *J. Biol. Chem.* 263, 7461-7464.
- Posfai, J., Bhagwat, A. S., Posfai, G., and Roberts, R. J. (1989). **Predictive motifs derived from cytosine methyltransferases.** *Nucleic Acids Res.* 17, 2421-2435.
- Posfai, J., Bhagwat, A. S., and Roberts, R. J. (1988). **Sequence motifs specific for cytosine methyltransferases.** *Gene* 74, 261-265.
- Powell, L. M., Dryden, D. T. F., Willcock, D. F., Pain, R. H., and Murray, N. E. (1993). **DNA recognition by the EcoK methyltransferase. The influence of DNA methylation and the cofactor S-adenosyl-L-methionine.** *J. Mol. Biol.* 234, 60-71.
- Powell, L. M., and Murray, N. E. (1995). **S-adenosyl methionine alters the DNA contacts of the EcoKI methyltransferase.** *Nucleic Acids Res.* 23, 967-974.
- Price, C., and Bickle, T. A. (1986). **A possible role for DNA restriction in bacterial evolution.** *Microbiol. Sci.* 3, 296-299.
- Price, C., Gubler, M., Braguglia, D., Meister, J., Tyndall, C., Piekarowicz, A., and Bickle, T. A. (1993). **Evolution of DNA sequence specificity in type IC restriction and modification systems.** *J. Cell Biochem. Suppl.* 17C, 151.
- Price, C., Lingner, J., Bickle, T. A., Firman, K., and Glover, S. W. (1989). **Basis for changes in DNA recognition by the EcoR124 and EcoR124/3 type I DNA restriction and modification enzymes.** *J. Mol. Biol.* 205, 115-125.
- Price, C., Shepherd, J. C. W., and Bickle, T. A. (1987). **DNA recognition by a new family of type I restriction enzymes: a unique relationship between two different DNA specificities.** *EMBO J.* 6, 1493-1497.
- Ralston, D. J., and Krueger, A. P. (1952). **Phage multiplication on two hosts, isolation and activity of variants of Staphylococcus phage P1.** *Proc. Soc. Exp. Biol. Med.* 80, 217-220.
- Ravetch, J. V., Horiuchi, K., and Zinder, N. D. (1978). **Nucleotide sequence of the recognition site for the restriction-modification enzyme of Escherichia coli B.** *Proc. Natl. Acad. Sci. U.S.A.* 75, 2266-2270.
- Reich, N. O., Maegley, K. A., Shoemaker, D. D., and Everett, E. (1991). **Structural and functional analysis of EcoRI DNA methyltransferase by proteolysis.** *Biochemistry* 30, 2940-2946.
- Reich, N. O., and Mashhoon, N. (1991). **Kinetic mechanism of the EcoRI DNA methyltransferase.** *Biochemistry* 30, 2933-2929.
- Reinisch, K. M. (1996). **The crystal structure of HaeIII methylase covalently complexed to DNA: An extrahelical cytosine and rearranged base pairing.** *Diss. Abstr.* 56, 3746-B.
- Reinisch, K. M., Chen, L., Verdine, G. L., and Lipscomb, W. N. (1995). **The crystal structure of HaeIII methyltransferase covalently complexed to DNA: an extrahelical cytosine and rearranged base pairing.** *Cell* 82, 143-153.
- Reiser, J., and Yuan, R. (1977). **Purification and properties of the P15 specific restriction endonuclease from Escherichia coli.** *J. Biol. Chem.* 252, 451-456.
- Renbaum, P., and Razin, A. (1992). **Mode of action of the Spiroplasma CpG methylase M.SssI.** *FEBS Lett.* 313, 243-247.
- Renbaum, P., and Razin, A. (1995). **Footprint analysis of M.SssI and M.HhaI methyltransferases reveals extensive interactions with the substrate DNA backbone.** *J. Mol. Biol.* 248, 19-26.



- Roberts, R. J. (1978a). **Restriction endonucleases**. In *Microbiology*, D. Schlessinger, ed. (Washington: American Society for Microbiology), pp. 5-9.
- Roberts, R. J. (1978b). **Restriction endonucleases: a new role in vivo?** *Nature* 271, 502.
- Roberts, R. J., and Macelis, D. (1996). **REBASE--restriction enzymes and methylases**. *Nucleic Acids Res.* 24, 223-235.
- Roberts, R. J., Myers, P. A., Morrison, A., and Murray, K. (1976). **A specific endonuclease from *Haemophilus haemolyticus***. *J. Mol. Biol.* 103, 199-208.
- Rosamond, J., Endlich, B., and Linn, S. (1979). **Electron microscopic studies of the mechanism of action of the restriction endonuclease of *Escherichia coli* B**. *J. Mol. Biol.* 129, 619-635.
- Rost, B., and Sander, C. (1993). **Prediction of protein structure at better than 70% accuracy**. *J. Mol. Biol.* 232, 584-599.
- Rost, B., and Sander, C. (1994). **Combining evolutionary information and neural networks to predict protein secondary structure**. *Proteins* 19.
- Rost, B., Sander, C., and Schneider, R. (1994). **PHD - a mail server for protein secondary structure prediction**. *CABIOS* 10, 53-60.
- Roulland-Dussoix, D., and Boyer, H. W. (1969a). **B Restriction endonuclease of *Escherichia coli***. *Bacteriol. Proc.* 69, 46.
- Roulland-Dussoix, D., and Boyer, H. W. (1969b). **The *Escherichia coli* B restriction endonuclease**. *Biochim. Biophys. Acta* 195, 219-229.
- Rountree, P. M. (1956). **Variations in a related series of staphylococcal bacteriophages**. *J. Gen. Microbiol.* 15, 266-279.
- Rubin, M. M., and Changeux, J.-P. (1966). **On the nature of allosteric transitions: implications of non-exclusive ligand binding**. *J. Mol. Biol.* 21, 264-274.
- Rubin, R. A., and Modrich, P. (1977). **EcoRI methylase physical and catalytic properties of the homogeneous enzyme**. *J. Biol. Chem.* 252, 7265-7272.
- Rutebuka, O. B., Lee, N. S., and Ryu, J. (1995). **DNA sequence analysis of the KpnAI restriction gene of *Klebsiella pneumoniae* M5a1**. *ASM Abstracts* 95, 522.
- Saha, S., and Rao, D. N. (1995). **ATP hydrolysis is required for DNA cleavage by EcoPI restriction enzyme**. *J. Mol. Biol.* 247, 559-567.
- Sain, B., and Murray, N. E. (1980). **The hsd (host specificity) genes of *E. coli* K12**. *Mol. Gen. Genet.* 180, 35-46.
- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). **Molecular Cloning: A Laboratory Manual**, Second Edition, Volume 1 (New York: Cold Spring Harbor Laboratory Press).
- Sander, C., and Schneider, R. (1991). **Database of Homology-Derived Structures and the Structural Meaning of Sequence Alignment**. *Proteins*, 56-68.
- Schaeffer, C. J., Huang, B., and Tsai, M. D. (1996). **A new class IIS zinc finger restriction enzyme with specificity for SP1 binding sites**. *Faseb J.* 10, A1241.
- Schluckebier, G., Labahn, J., Granzin, J., and Saenger, W. (1995a). **Structure and function of the adenine specific DNA methyltransferase M.TaqI**. *J. Biomol. Struct. Dyn.* 12, A206.
- Schluckebier, G., O'Gara, M., Saenger, W., and Cheng, X. (1995b). **Universal catalytic domain structure of AdoMet-dependent methyltransferases**. *J. Mol. Biol.* 247, 16-20.
- Schulz, G. (1988). **A critical evaluation of methods for prediction of protein secondary structures**. *Annu. Rev. Biophys. Biophys. Chem.* 17, 1-21.



- Selent, U., Ruter, T., Kohler, E., Liedtke, M., Thielking, V., Alves, J., Oelgeschlager, T., Wolfes, H., Peters, F., and Pingoud, A. (1992). **A site-directed mutagenesis study to identify amino acid residues involved in the catalytic function of the restriction endonuclease EcoRV.** *Biochemistry* 31, 4808-4815.
- Sharp, P. M. (1986). **Molecular evolution of bacteriophages: Evidence of selection against the recognition sites of host restriction enzymes.** *Mol. Biol. Evol.* 3, 75-83.
- Skowron, P., Kaczorowski, T., Tucholski, J., and Podhajska, A. J. (1993). **Atypical DNA-binding properties of class-IIIS restriction endonucleases: evidence for recognition of the cognate sequence by a FokI monomer.** *Gene* 125, 1-10.
- Skrzypek, E., and Piekarowicz, A. (1989). **The EcoDXX1 restriction and modification system: cloning the genes and homology to type I restriction and modification systems.** *Plasmid* 21, 195-204.
- Slavik, J. (1982). **Anilino-naphthalene sulfonate as a probe of membrane composition and function.** *Biochim. Biophys. Acta* 694, 1-25.
- Smith, H. O., and Nathans, D. (1973). **A suggested nomenclature for bacterial host modification and restriction systems and their enzymes.** *J. Mol. Biol.* 81, 419-423.
- Smith, H. O., and Wilcox, K. W. (1970). **A restriction enzyme from Hemophilus influenzae. I. Purification and general properties.** *J. Mol. Biol.* 51, 379-391.
- Smith, S. S., Kaplan, B. E., Sowers, L. C., and Newman, E. M. (1992). **Mechanism of human methyl-directed DNA methyltransferase and the fidelity of cytosine methylation.** *Proc. Natl. Acad. Sci. U.S.A.* 89, 4744-4748.
- Som, S., Bhagwat, A. S., and Friedman, S. (1987). **Nucleotide sequence and expression of the gene encoding the EcoRII modification enzyme.** *Nucleic Acids Res.* 15, 313-332.
- Spoerel, N., Herrlich, P., and Bickle, T. A. (1979). **A novel bacteriophage defence mechanism: the anti-restriction protein.** *Nature* 278, 30-34.
- Studier, F. W., and Moffat, B. A. (1986). **Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes.** *J. Mol. Biol.* 189, 113.
- Sturrock, S. S., and Dryden, D. T. F. (1997). **A prediction of the amino acids and structures involved in DNA recognition by type I DNA restriction and modification enzymes.** *Nucleic Acids Res.* 25, 3408-3414.
- Subramanya, H. S., Bird, L. E., Brannigan, J. A., and Wigley, D. B. (1996). **Crystal structure of a DExx box DNA helicase.** *Nature* 384, 379-383.
- Suck, D. (1994). **Flip out and modify.** *Curr. Biol.* 4, 252-255.
- Surby, M. A., and Reich, N. O. (1992). **The role of nonspecific DNA in the site-location kinetics of the EcoRI DNA methyltransferase.** *Biochemistry* 31, 2208.
- Surby, M. A., and Reich, N. O. (1996). **Facilitated diffusion of the EcoRI DNA methyltransferase is described by a novel mechanism.** *Biochemistry* 35, 2209-2217.
- Suri, B., and Bickle, T. A. (1985). **EcoA: The first member of a new family of Type I restriction modification systems - Gene organization and enzymatic activities.** *J. Mol. Biol.* 186, 77-85.
- Suri, B., Nagaraja, V., and Bickle, T. A. (1984). **Bacterial DNA modification.** *Curr. Top. Microbiol. Immunol.* 108, 1-9.
- Suri, B., Shepherd, J. C. W., and Bickle, T. A. (1984). **The EcoA restriction and modification system of Escherichia coli 15T-: enzyme structure and DNA recognition sequence.** *EMBO J.* 3, 575-579.



- Szilak, L., Der, A., Deak, F., and Venetianer, P. (1993). **Kinetic characterization of the EcaI methyltransferase.** *Eur. J. Biochem.* **218**, 727-733.
- Szybalski, W., Kim, S. C., Hasan, N., and Podhajska, A. J. (1991). **Class-IIS restriction enzymes - a review.** *Gene* **100**, 13-26.
- Tabor, S. (1990). **Expression using the T7 RNA Polymerase/Promoter System.** In *Protocols in Molecular Biology*, F. A. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith and K. Struhl, eds. (New York: Greene Publishing and Wiley-Interscience), pp. 16.2.1-16.2.11.
- Takano, K., Nakamura, T., and Sekiguchi, M. (1991). **Roles of two types of O6-methylguanine-DNA methyltransferases in DNA repair.** *Mutat. Res.* **254**, 37-44.
- Takano, T., Watanabe, T., and Fukasawa, T. (1966). **Specific inactivation of infectious lambda DNA by sonicates of restrictive bacteria with R factors.** *Biochem. Biophys. Res. Commun.* **25**, 192-204.
- Taylor, I., Patel, J., Firman, K., and Kneale, G. (1992). **Purification and biochemical characterisation of the EcoR124 type I modification methylase.** *Nucleic Acids Res.* **20**, 179-186.
- Taylor, I., Watts, D., and Kneale, G. (1993). **Substrate recognition and selectivity in the type IC DNA modification methylase M.EcoR124I.** *Nucleic Acids Res.* **21**, 4929-4935.
- Taylor, I. A. (1992). **Biochemical characterisation of the EcoR124 type I DNA modification methylase.** Thesis, submitted to , University of Portsmouth, Portsmouth.
- Taylor, I. A., Davis, K. G., Watts, D., and Kneale, G. G. (1994). **DNA binding induces a major structural transition in a type I methyltransferase.** *EMBO J.* **13**, 5772-5778.
- Thorpe, P. H., Ternent, D., and Murray, N. E. (1997). **The specificity of StySKI, a type I restriction enzyme, implies a structure with rotational symmetry.** *Nucleic Acids Res.* **25**, 1694-1700.
- Titheradge, A. J. B., Ternent, D., and Murray, N. E. (1996). **A third family of allelic hsd genes in Salmonella enterica: sequence comparisons with related proteins identify conserved regions implicated in restriction of DNA.** *Mol. Microbiol.* **22**, 437-447.
- Topal, M. D., and Conrad, M. (1993). **Changing endonuclease EcoRII Tyr308 to Phe abolishes cleavage but not recognition: possible homology with the Int-family of recombinases.** *Nucleic Acids Res.* **21**, 2599-2603.
- Trautner, T. A., Balganesch, T., Wilke, K., Noyer-Weidner, M., Rauhut, E., Lauster, R., Behrens, B., and Pawlek, B. (1988a). **Organization of target-recognizing domains in the multispecific DNA (cytosine-5) methyltransferases of Bacillus subtilis phages SPR and Phi-3T.** *Gene* **74**, 267.
- Trautner, T. A., Balganesch, T. S., and Pawlek, B. (1988b). **Chimeric multispecific DNA methyltransferases with novel combinations of target recognition.** *Nucleic Acids Res.* **16**, 6649-6658.
- Tyndall, C., Meister, J., and Bickle, T. A. (1994). **The Escherichia coli prr region encodes a functional type IC DNA restriction system closely integrated with an anticodon nuclease gene.** *J. Mol. Biol.* **237**, 266-274.
- Valinluck, B., Lee, N. S., and Ryu, J. (1995). **A new restriction-modification system, KpnBI, recognized in Klebsiella pneumoniae.** *Gene* **167**, 59-62.
- Verdine, G. (1994). **The flip side of DNA methylation.** *Cell* **76**, 197-200.
- Vidgren, J., Svensson, L. A., and Liljas, A. (1994). **Crystal structure of catechol O-methyltransferase.** *Nature* **368**, 354-358.
- Vipond, I. B., Baldwin, G. S., and Halford, S. E. (1995). **Divalent metal ions at the active sites of the EcoRV and EcoRI restriction endonucleases.** *Biochemistry* **34**, 697-704.



- Vogelstein, B., and Gillespie, D. (1979). **Preparative and analytical purification of DNA from agarose.** *Proc. Natl. Acad. Sci. USA* 76, 615-619.
- Wah, D. A., Hirsch, J. A., Dorner, L. F., Schildkraut, I., and Aggarwal, A. K. (1997). **Structure of the multimodular endonuclease FokI bound to DNA.** *Nature* 388, 97-100.
- Wang, R. Y. H., Kuo, K. C., Gehrke, C. W., Huang, L. H., and Ehrlich, M. (1982). **Heat- and alkali-induced deamination of 5-methylcytosine and cytosine residues in DNA.** *Biochim. Biophys. Acta* 697, 371-377.
- Watanabe, T., Nishida, H., Ogata, C., Arai, T., and Sato, S. (1964). **Episome-mediated transfer of drug resistance in enterobacteriaceae. VII. Two types of naturally occurring R factors.** *J. Bacteriol.* 88, 716-726.
- Waugh, D. S., and Sauer, R. T. (1993). **Single amino acid substitutions uncouple the DNA binding and strand scission activities of Fok I endonuclease.** *Proc. Natl. Acad. Sci. U.S.A.* 90, 9596-9600.
- Waugh, D. S., and Sauer, R. T. (1994). **A novel class of FokI restriction endonuclease mutants that cleave hemi- methylated substrates.** *J. Biol. Chem.* 269, 12298-12303.
- Weiserova, M., Janscak, P., Benada, O., Hubacek, J., Vitaly, E. A., Glover, S. W., and Firman, K. (1993). **Cloning, production and characterization of wild type and mutant forms of the R.EcoK endonucleases.** *Nucleic Acids Res.* 21, 373-379.
- Wilke, K., Rauhut, E., Noyer-Weidner, M., Lauster, R., Pawlek, B., Behrens, B., and Trautner, T. A. (1988). **Sequential order of target-recognizing domains in multispecific DNA-methyltransferases.** *EMBO J.* 7, 2601-2609.
- Willcock, D. F., Dryden, D. T. F., and Murray, N. E. (1994). **A mutational analysis of the two motifs common to adenine methyltransferases.** *EMBO J.* 13, 3902-3908.
- Wilson, G. G. (1992). **Amino acid sequence arrangements of DNA-methyltransferases.** *Meth. Enzymol.* 216, 259-279.
- Winkler, F. K. (1994). **DNA totally flipped-out by methylase.** *Structure* 2, 79-83.
- Winkler, F. K., Banner, D. W., Oefner, C., Tsernoglou, D., Brown, R. S., Heathman, S. P., Bryan, R. K., Martin, P. D., Petratos, K., and Wilson, K. S. (1993). **The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments.** *EMBO J.* 12, 1781-1795.
- Winter, M. (1992). **Reinigung und Charakterisierung der M-Untereinheit des Restriktions-/Modificationssystems von E. coli K-12.** Thesis, submitted to the Fakultät für Biologie, Julius-Maximilians-Universität, Würzburg.
- Woelcke, J., and Weinhold, E. (1995). **Substrate specificity of the DNA-methyltransferase from Thermus aquaticus: influence of the 3'-neighbor base.** *Biol. Chem. Hoppe-Seyler* 376, S169.
- Wood, W. B. (1965). **Mutations in E. coli affecting the host-controlled modification bacteriophage lambda.** *Path. Microbiol.* 28, 73-76.
- Wu, J. C., and Santi, D. V. (1987). **Kinetic and catalytic mechanism of HhaI methyltransferase.** *J. Biol. Chem.* 262, 4778-4786.
- Wu, J. C., and Santi, D. V. (1988). **High level expression and purification of HhaI methyltransferase.** *Nucleic Acids Res.* 16, 703-717.
- Wu, J. C., Santi, D. V., and Razin, A. (1985). **On the mechanism and inhibition of DNA cytosine methyltransferases.** *Biochemistry and Biology of DNA Methylation* Alan R. Liss, Cantoni, G.L., 119-129.
- Wyatt, G. R. (1950). **Occurrence of 5-Methyl-Cytosine in Nucleic Acids.** *Nature* 166, 237-238.



- Wyszynski, M. W., Gabbara, S., Kubareva, E. A., Romanova, E. A., Oretskaya, T. S., Gromova, E. S., Shabarova, Z. A., and Bhagwat, A. S. (1993). The cysteine conserved among DNA cytosine methylases is required for methyl transfer, but not for specific DNA binding. *Nucleic Acids Res.* 21, 295-301.
- Xu, G., Willert, J., Kapfer, W., and Trautner, T. A. (1995). BsuCI, a type-I restriction-modification system in *Bacillus subtilis*. *Gene* 157, 59.
- Yang, A. S., Shen, J.-C., Zingg, J.-M., Mi, S., and Jones, P. A. (1995). HhaI and HpaII DNA methyltransferases bind DNA mismatches, methylate uracil and block DNA repair. *Nucleic Acids Res.* 23, 1380-1387.
- Yonezawa, A., and Sugiura, Y. (1994). DNA binding mode of class-IIS restriction endonuclease FokI revealed by DNA footprinting analysis. *Biochim. Biophys. Acta* 1219, 369-379.
- Yoon, H., Suh, H., Kim, K., Han, M. H., and Yoo, O. J. (1985). The specificity and catalytic properties of AluI methylase. *Korean Biochem. J.* 18, 88-93.
- Yoshimori, R. N. (1971). A genetic and biochemical analysis of the restriction and modification of DNA by resistance transfer factors. Thesis, submitted to , University of California, San Francisco, USA.
- Yuan, R., Bickle, T. A., Ebbers, W., and Brack, C. (1975). Multiple steps in DNA recognition by restriction endonuclease from *E. coli* K. *Nature* 256, 556-560.
- Yuan, R., Hamilton, D. L., Hadi, S. M., and Bickle, T. A. (1980). Role of ATP in the cleavage mechanism of the EcoP15 restriction endonuclease. *J. Mol. Biol.* 144, 501-519.
- Yuan, R., and Reiser, J. (1978). Steps in the reaction mechanism of the *Escherichia coli* plasmid P15-specific restriction endonuclease. *J. Mol. Biol.* 122, 433-445.
- Zabeau, M., Friedman, S., Van Montagu, M., and Schell, J. (1980). The *ral* gene of phage lambda: I. Identification of a non-essential gene that modulates restriction and modification in *E. coli*. *Mol. Gen. Genet.* 179, 63-73.
- Zinder, N. D. (1960). Hybrids of *Escherichia* and *Salmonella*. *Science* 131, 813-815.